DOCUMENT RESUME

ED 327 726 CE 056 816

AUTHOR Gilmore, Alison; Wagner, Graham
TITLE The Readability of Trade Examinations.
INSTITUTION New Zealand Council for Educational Research, Wellington.
REPORT NO ISBN-0-908567-49-9
PUB DATE 85
NOTE 97p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Achievement Tests; Agricultural Education; *Apprenticeships; Auto Mechanics; Cloze Procedure; Cosmetology; *Difficulty Level; Educational Research; Electricians; Farmers; Foreign Countries; Meat Packing Industry; Occupational Home Economics; Postsecondary Education; *Readability; Reading Ability; *Reading Achievement; Reading Comprehension; Secondary Education; Student Certification; *Trade and Industrial Education; Vocational Education; Writing Ability; Writing Achievement; Writing Evaluation
IDENTIFIERS *New Zealand

ABSTRACT

A 1982 study assessed the comparative reading comprehension levels and writing ability performance of Trades Certification Board (TCB) apprentices in five trades in New Zealand. Test data were collected for 1,389 apprentices and a comparison group of 361 secondary school students. The cloze procedure format was adopted for assessing apprentices' ability to gain information from both general reading material and trade reading material. Overall, for all trades except hairdressing, the ability of apprentices to read extracts from their respective trade materials was lower than their ability to tackle general reading materials. The electrical, farming, and auto engineering course trade materials had significantly lower readability than meat retailing materials, which had significantly lower readability than hairdressing materials. Reading and writing abilities were moderately related. A readability analysis of the 1982 First Qualifying TCB examination papers used comparable samples of 149 apprentices selected from the 1983 intakes. The cloze format was again used. Auto engineering and meat retailing groups had trouble reading their examination papers independently. Examination papers appeared to be very similar in readability, but in practice this penalized some groups. The report includes 43 references and appendixes containing study materials, including information on the cloze procedure and sample cloze tests. (YLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED327726

THE READABILITY OF TRADE EXAMINATIONS

Alison Gilmore and Graham Wagner

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K A Pickens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

New Zealand Council for Educational Research
1985

CE 056 816

CONTENTS

Acknowledgements

1.0	Introduction	1
	1.1 The Problem	2
	1.2 Defining Reading Comprehension and Readability	5
	1.3 The Cloze Procedure	6
	1.4 Objectives of the Study	7
2.0	Part I: Methodology	8
	2.1 Sample	8
	2.2 The Test of Reading Comprehension	8
	2.3 Research Design: Multiple Matrix Sampling (MMS)	11
	2.4 The Test Booklets	12
	2.5 Writing	12
	2.6 Administration of the Tests	12
	2.7 Analyses of the Data	13
3.0	Part I: Results	14
	3.1 Sample Characteristics	14
	3.2 Performance on Reading Comprehension Tests	18
	3.3 Reading Comprehension Performance Based on Criterion Scores	19
	3.4 Reading Comprehension Scores Based on Means and Standard Deviations	27
	3.5 Consistency of Reading Performance in the General and Trade Sections	32
	3.6 Writing Ability	33
	3.7 Relationship between Reading Comprehension and Writing Ability	34
	3.8 Reading Comprehension, Writing and First Qualifying Examination Performance	35

4.0	Part II: Methodology	37
	4.1 Sample	37
	4.2 The Test Booklets	37
	4.3 Administration of the Tests	38
	4.4 Analysis of the Data	38
5.0	Part II: Results	38
	5.1 Examination Cloze Performance based on Criterion Scores	38
	5.2 Examination Readability based on Means and Standard Deviations	41
6.0	Conclusions	45
	6.1 Comprehension of General and Trade Reading Materials	45
	6.2 Relationship between Reading Comprehension, Writing and Examination Performance	46
	6.3 Examination Readability	46
7.0	Recommendations	47
	References	49
	Appendices	53

ACKNOWLEDGEMENTS

We would like to thank the many tutors and testing coordinators who helped us with this survey. In particular we would like to acknowledge the full cooperation and support given to us by the Principals, staff and apprentices of the following institutions:

Manakau Technical Institute
Waikato Technical Institute
Bay of Plenty Community College
Taranaki Polytechnic
Wellington Polytechnic
Petone Technical Institute
Christchurch Polytechnic
Southland Community College
Smedley Farm Training Centre
Taratahi Farm Training Centre
Flock House
Telford Farm Training Centre

Finally, but not least, we sincerely appreciate the advice and competent assistance from Ian Livingstone, David Philips, Felicity O'Connor and Anna Aitken all of NZCER.

Alison Gilmore
Graham Wagner

1.0 INTRODUCTION

Early in 1981 the New Zealand Council for Educational Research (NZCER) was invited to collaborate with the Trades Certification Board (TCB) over the procedures to be adopted in future for assisting students with learning disabilities who are required to sit TCB examinations.[1] It seemed that a significant number of examination candidates had some reading problems. That this had become an important issue is shown by the involvement of the Chief Ombudsman who, at the time, was investigating the subject of reader/writer assistance for some students sitting TCB examinations.[2]

Subsequently, a small sub-committee, set up by TCB to consider the provision of reading and writing assistance to those with learning disabilities who were sitting TCB examinations, resolved that "the board ask NZCER to survey from selected trades the reading and writing ability of typical apprentices and to examine the reading levels of the existing examination papers for those trades". It was also resolved that "the NZCER assessment should have regard for the fact that the vocabulary used, while apparently of a specialist nature to the layman, is known as normal to the tradesman". The following trades were selected for the survey: Farming, Automotive Engineering, Electrical, Meat Retailing, and Ladies Hairdressing.[3]

Later, a meeting was held between a sub-committee of TCB and NZCER to discuss a feasibility study of the problem to clarify the issues concerned in the proposed survey. NZCER then contacted TCB and indicated that a readability survey was a viable proposition and that, subject to Council approval, NZCER staff, Dr. Gilmore and Mr Wagner, would be able to start the survey in 1982. In reply, TCB resolved that NZCER be asked to:

(1) assess the reading comprehension of students in the five trade areas of Farming, Automotive Engineering, Electrical, Meat Retailing, and Hairdressing; and

---

1 Letter to NZCER's Director, John Watson, from the Assistant Secretary (Tertiary), Department of Education, Mr Kings

2 After his first contact with NZCER, to see if the Council was involved in any projects of this kind, he indicated that TCB would approach NZCER for assistance with the reader/writer issue.

3 TCB minutes of a special sub-committee, 7 May 1981.

(2) determine the readability levels of TCB examination papers in these subjects.

It was also understood that an assessment of writing skills would be included in the survey, even though no mention was made at this stage about the writing ability of students.

Once the project was approved by the Council of NZCER, planning began in March 1982. Nationally representative samples of school pupils and apprentices were tested. By late November 1982 the testing of reading comprehension and writing abilities in the five apprentice groups had been completed (Stage I). The second stage of the research was undertaken in late 1983 (Stage II).

Data from Stage I results were processed in the first 6 months of 1983. Stage II started in August 1983 and the results were processed by the end of November. A resume of the main findings of the study was forwarded to TCB in May, 1984, prior to release of this report.

Structure of Report

The methodology and results sections of this report are divided into two parts. Part I discusses the reading comprehension and writing abilities of the 1982 samples of first-year apprentices in Farming, Meat Retailing, Electrical, Hairdressing and Automotive Engineering trades. Part II covers the performance of a comparable group of 1983 first-year apprentices in Meat Retailing, Electrical, Hairdressing and Automotive Engineering on readability tests of the 1982 TCB First Qualifying trade examination papers.

The findings of Parts I and II are drawn together in the final section of the report.

1.1 The Problem

Of the 50,000 or more students enrolling for TCB examinations in 1982, a small but significant proportion are likely to have encountered learning problems as they study towards their trade certificates. The small amount of research undertaken on the learning difficulties of such students (Callaghan, 1981; Stevens, 1981) indicates, that at least in the area of reading comprehension, there are real teaching difficulties

facing the tutors of technical institutes, community colleges, and other trade training institutions (Wagner, 1980; Marriot, 1981; and Nicholson, 1981). The first two studies actually show that, in two New Zealand institutes, larger numbers of students than expected had severe reading problems when tested by cloze tests of general reading comprehension on trade-related materials.

Overseas research on literacy amongst apprentices is as sparse as it is in New Zealand, but one study in Australia draws attention to the difficulties apprentices face in reading trade materials without the assistance of prior technical content knowledge and adequate language processing abilities (Mealyea,1981). Similar investigations have been reported in the USA (Karnes and Gunn, 1976; Sloan and Pearson, 1978; Thornton, 1977 and 1979) and recently in New Zealand (Marriot, 1981; Holmes, 1981). The point is made by Thornton (1979) that most research on the ease of reading of occupational literature has been directed towards the reading of text-books and an assessment of basic literacy needs. In his view, what is needed is "a study which examines a representative sample of all occupational specialists to determine what the requirements are in order to read employment-related literature. The reading difficulty of the literature for carpentry, dental assistant, home economics, practical nursing, and all the rest is simply not known" (p.15). The same need applies in New Zealand, although steps are now underway at Christchurch Polytechnic to investigate this matter and to develop methods of making trade materials more learnable and readable (Elley and Marriot, 1983).

Similarly, the readability of trade examinations has rarely been investigated. To our knowledge no studies have been undertaken which address the question "How readable are the trade examinations?" In practice, examinations are thought to be valid and reliable if they are properly constructed, well moderated, and students' performances indicate that the majority have understood those questions they have answered correctly. It follows that if the the majority have the right answers (at least on a wide range of questions) then the reason some students do not do well is due solely to their lack of technical ability (i.e. content knowledge). Readability is not usually seen as a barrier to accurate student expression. For example, most examiners fail to check that the readability levels of their examination papers are appropriate

for the group being tested, firstly, because it is believed that the examiner is able to judge whether the examination is readable or not (we now know this to be a fallacy) and, secondly, measuring the readability of examinations is no easy matter. To our knowledge there is only one readability index designed for standardised tests, but this index is based upon an out-of-date edition of the Thorndike Junior Century Dictionary (Forbes and Cottle,1953) and, therefore is of little practical use.

What is needed, then, is a method of assessing whether trade apprentices can read with understanding the questions in their trade examinations, and a way to determine the extent to which other factors related to the examination itself (e.g. familiarity with examination format, question type, language, test anxiety, lack of content knowledge, etc.) influence examination performance. While all of these factors are significant hurdles to overcome for the apprentice, or anyone sitting formal examinations for that matter, the fundamental one is reading literacy.

In addition to reading skills, trade apprentices require writing skills in order to communicate to an examiner that they have read and understood the questions, and that the answers they write (here we are referring to open-ended questions) represent accurately the amount of knowledge they possess. The relationship between reading and writing ability is thought to be complementary (Parker and Davis, 1983 , p.55); thus, if one is deficient, the other is likely to be deficient also. However, while it is known that good writers are also good readers (a close relationship exists) there is some evidence that the reverse does not always apply (Haynes, 1978). Hence good readers may not always do well in their written examinations because of their lack of writing skill. As trade examinations require apprentices to write coherently, a study of writing literacy among the apprentices is germane to a study of trade reading literacy, especially as it is expected that relevant knowledge must be operationalised (i.e. translated from theory to practice and explained in the apprentice's own words).

In New Zealand in recent years there has been a vigorous debate at times about the literacy levels of school pupils, trade apprentices and adults. Two studies of third and fourth form New Zealand high school

students some years ago indicated that more than half were unable to understand their basic subject textbooks (Boag, 1975; Elley,1976). As many trade apprentices are recruited from these two levels it is likely that the reading difficulties experienced by lower secondary school students are also encountered by numbers of apprentices, especially when reading their trade materials. Callaghan (1981) reports that a literacy problem remains widespread at the post-secondary trade level despite special efforts to produce 'readable' trade and professional study materials which often replace the traditional textbooks.

A similar situation arises in the field of adult education. The growth in adult reading assistance schemes reveals that significant numbers of New Zealand adults can be described as functionally illiterate (Sommerville and Chapman,1979), if the UNESCO reading age of 13 years is an acceptable criterion of functional literacy. Thus the reading and writing skills of trade apprentices need to be studied to ensure that their technical training is beneficial to them and to the country as a whole.

1.2 Defining Reading Comprehension and Readability

At this point, it is essential to clarify various terms and procedures used to measure reading comprehension and readability. Scholars have had great difficulty defining comprehension, while to define reading comprehension as the ability of the reader to comprehend or understand what she/he has read is very uninformative. No one can really describe the *process* of comprehension. This difficulty is avoided if comprehension is defined in terms of the output of the reading act. Bormuth (1971) and others have shown that if an individual can 'gain information' as a result of reading verbal materials, then comprehension has occurred. Bormuth describes the output of the reading act in this way: "Comprehension ability is thought to be a set of generalised acquisition skills which permit people to acquire and exhibit *information gained* as a consequence of reading printed material".

Readability, on the other hand, is defined as those aspects of a text which make it easy (or hard) for a reader to understand, e.g., legibility, illustration, colour, vocabulary, conceptual difficulty, syntax and organisation of content. In measuring text difficulty, factors

relating to the reader, such as reading ability, reading age and motivation also have to be considered.

Measuring readability (i.e. text difficulty) or reading comprehension (i.e. the reader's ability to gain information from the text) has been simplified by the creation of the cloze procedure (Taylor, 1953, 1954, 1956) and its development, especially by Bormuth (1962, 1965, 1967, 1968, 1969, 1970, 1971). The cloze procedure, created as a measure of text readability, is also regarded as the best means to assess the reading comprehension abilities of readers.

1.3 The Cloze Procedure[4]

Elley (1976) has written that "In its simplest form, the cloze procedure is a method of testing reading. Words are deleted from a prose passage, either randomly or according to rule, and the reader is required to replace them correctly, using the context as a guide. To do this, he goes through a process ... called 'clozure'. He closes the gap in the information: hence the name 'cloze'"(p.35). Cloze tests measure the readers' knowledge of the subject, their understanding of language structures and their ability to draw inferences. Elley (1978) concludes his review of the technique by saying that there are good reasons for accepting it"... as a valid and useful test of reading comprehension"(p.59). An example of one of the cloze tests used in this study, with the accompanying introduction is contained in Appendix I.5.

The cloze has been demonstrated to be extremely versatile in its application to assessing the readability of a wide variety of written material. The technique has been applied to prose, poetry, references, mathematics, titles, directions, advertisements, and instructions (Hanson and Hesse, 1974). The cloze is the most useful procedure available for determining the readability of TCB trade examinations since it can provide a *direct* measure of the apprentice's ability to comprehend the examination paper. With reference to standard cut-off points, it can be used to identify the approximate proportions of apprentices experiencing varying degrees of difficulty in reading and understanding their trade-related written materials.

There is now a substantial body of research on the way that the cloze can be used for measuring reading comprehension levels and the readability of text. Bormuth (1975) lays down the standard procedure for

---

[4] For a more detailed discussion, see Appendix I.

developing and administering cloze tests. The interpretation and use of cloze test scores in both reading comprehension and readability assessment are discussed in Appendix I.3.

1.4 Objectives of the Study

The aims of this study are broadly sketched in the Introduction. Specifically they are:

For TCB apprentices in Electrical, Hairdressing, Automotive Engineering, Farming and Meat Retailing trades, to determine comparative

(a) reading comprehension levels;

(b) writing ability performance;

(c) readability levels on end-of-year trade examination papers.

In achieving these aims, the study addressed a number of questions:

1. What proportion of apprentices within each trade is experiencing difficulties with reading 'General' and 'Trade' materials?
2. Are there differences in the reading comprehension skills among apprentices in Hairdressing, Meat Retailing, Electrical, Farming and Automotive Engineering?
3. Within each trade group, are there differences in the apprentices' ability to comprehend 'General' and 'Trade' materials?
4. Are there differences among the apprentice groups in their ability to write?
5. What proportion of apprentices in each trade group is experiencing difficulty in reading their end-of-year examination papers?
6. Are there differences among the apprentice groups in their ability to read their end-of-year examinations?

PART I

2.0 METHODOLOGY

2.1 Sample

Five populations of trade apprentices were of primary interest in this study: Automotive Engineering, Electrical, Farming, Ladies Hairdressing and Meat Retailing. These populations were further defined as comprising apprentices in their first year of training who were attempting the Trades Certification Board First Qualifying examination in December, 1982.

Seven technical institutes were selected to be broadly representative of those throughout New Zealand by location (North Island, South Island), size (large urban, urban) and type. Not all of the five trade areas of interest were available at each institute selected but , for each trade, a range of institutes was represented. Additional first-year Farming cadets were available from four further institutes and these were also included in the study. (This provided the national cohort of first-year Farming cadets). Apprentices studying with the Technical Correspondence Institute were also represented in the samples. The total number of apprentices tested in each trade area was: Meat Retailing, 209; Electrical, 241; Ladies Hairdressing, 321; Automotive Engineering, 334; and Farming, 384.

In order to compare these groups with a sample of 'young adults' from the general population, samples of Form 4 and Form 5 secondary school students (from 14 to 17 years old) were drawn from 13 schools. These schools were nationally representative by education region and locality (large urban, urban, town and rural). In total, test data for 361 secondary school students (182 males and 179 females, 197 Form 4 and 164 Form 5) were collected.

2.2 The Test of Reading Comprehension

The major focus of the study was an evaluation of the reading comprehension ability (that is, the ability to gain information from textual material) of apprentices in the five nominated trade areas. As the cloze procedure has been shown to be a valid and reliable measure of reading comprehension or 'information gain', this format was adopted for assessing the amount of information an individual was able to absorb from two categories of reading materials:

(1) general material that 'young adults' may be expected to encounter in their daily living; and

(2) reading material specific to particular trades, which apprentices might be expected to read during the course of their study or work.

Identifying the Reading Materials

General Reading Material

An extensive survey of literacy (Hanson and Hesse, 1974) identified over 60 'categories' of general reading material, classified into 10 major 'domains': recreational literature, automobile, citizenship, leisure time activities, occupational, safety, reference, consumer material, school job-related and text-books. Examples of 'categories' within the domain of consumer materials were: banking promotional literature, advertisements, contractual agreements, junk mail, messages on packages and recipes. For this study a sample of 'categories' from each domain was chosen, except for text-book material, as this was covered by the trade-specific reading material selected from course work.

TABLE 2.1 Domains and Categories of Reading Materials

Recreational Literature

1. Fiction
2. Non-fiction

Automobile

3. New Zealand Transport Department Road Code

Citizenship

4. Newspaper article

Leisure Time Activities

5. Travel brochure

Occupational

6. Prospective career promotional literature

Safety

7. New Zealand Health Department literature

Reference

8. Reference books
9. Telephone directory

Consumer

10. Banking promotional literature
11. Advertisements

School Job-Related

12. Student Prospectus
13. Contract of apprenticeship

Table 2.1 presents a list of the categories and domains sampled. A number of passages from sources representing this range of reading material were identified, and submitted to five specialists in the area of reading in technical institutes and adult literacy programmes for examination. Each person was asked to comment on how representative each passage was of the reading material that young adults are likely to encounter in daily life, to select the 13 most appropriate passages, and to rank order them in terms of difficulty. For reasons which will be given later, 13 passages of about 150 - 170 words were chosen for the test, and the order of difficulty of the passages was established by combining the rank orders on readability from the five counsultants for each test passage, and checking this by use of the Noun Frequency Count (see Elley, 1975, for a description of this method).

Trade Reading Materials

From each technical institute in the study, including the Technical Correspondence Institute (TCI), lists and examples of textbooks, work manuals, job sheets, trade regulations and course materials prepared by tutors were gathered for each trade area. A selection of sources, representing a cross-section of the materials supplied by the institute, was sampled. From these, extracts were identified for scrutiny by TCI staff. The consultants were requested to select the seven most appropriate passages, representing the types, level and content of material first-year apprentices were likely to encounter during the course of their work or study. These passages were also ranked in order of difficulty.

'Clozing' the Test Passages

Thirteen general reading passages and 35 trade passages (seven for each trade area) were prepared for the tests. Each passage had a title (to provide a context) and was a 150 - 170 word sequence of writing, beginning with a new paragraph, and ending with a sentence. Each passage was then 'clozed' to generate 20 test items (the procedure

adopted is outlined in Appendix I.4)

2.3 Research Design: Multiple Matrix Sampling (MMS)

The primary interest of this study was the performance of different *groups* of students rather than the performance of individuals. A relatively new technique, Multiple Matrix Sampling (MMS) was used to allow a wide range of materials to be sampled without overburdening individual apprentices. Multiple Matrix Sampling requires random samples of individuals to respond to randomly selected subsets of test items, so that the different groups answer different sets of questions. The performances of each subgroup on each subtest are combined, and an estimate made of how the group as a whole would perform on *all* the test passages. Much evidence (e.g. Lord, 1962; Plumlee, 1964; Shoemaker, 1973) has shown that the use of several samples of test items from a pool of items provides very acceptable estimates of such population parameters as the mean, standard deviation and frequency distribution where the statistics from subtest samples are appropriately combined.

In order to adopt the best MMS research design, several factors were taken into account, such as the need to obtain measures of reading ability (general and trade) from as broad a range of reading materials as possible, the 50 minutes available for testing, the number of blanks that the test takers could be expected to respond to without fatigue or boredom influencing test performance (about 160), and the total number of apprentices available for testing (from 240 to 400). A sample size of approximately 60, with 80 items per test, provided a standard error of less than 3 which was found acceptable for the level of analysis to be performed (further details on the sampling plan are contained in Appendix II).

Four parallel subtests of general reading, plus two parallel subtests of trade reading were required. Each section needed four passages, one of which was common to each of the parallel subtests so that checks on the equivalence of the samples could be made. A total of 13 general reading passages and seven trade reading passages was required for each trade. The test passages were arranged in order of difficulty and grouped into four strata of difficulty for the general section and two strata for the trade section. Test passages were then randomly assigned

from each stratum to one of the respective subtests.

The four test booklets for each trade were then distributed randomly so that the questions in each test booklet were attempted by about one quarter of the apprentices taking the test.

2.4 The Test Booklets

The test booklets for trade apprentices consisted of two sections on reading comprehension. In the first (Section A), the material was general, with four parallel subtests, and in the second (Section B), the material was trade specific with two parallel subtests. In compiling the two sections, procedures were adopted to minimise any possible practice effects by systematically distributing passages to create 16 forms of test booklets for each trade. In total, 80 different forms of test booklets were compiled over the five trade areas (see Appendix III for details on how possible serial effects were guarded against).

An important component of the MMS procedures is that the test booklets are randomly distributed to test takers. To ensure this was done, the collated test booklets were arranged in cycles of 16 so that within each cycle one of each version of the test was present. Tests were then despatched to the technical institutes for administration. The test booklets for the secondary school sample contained only the general reading comprehension section.

2.5 Writing

Also of concern in this study was an examination of the writing ability of the apprentice groups. A sample of each apprentice's writing was obtained by asking them to write 10 - 15 lines on the topic "Why I became an apprentice", after they had completed the cloze reading sections. Ten minutes were allocated for this. The secondary school sample, naturally, was not required to complete this writing exercise. (Examples of the test booklets are presented in Appendix I.5).

2.6 Administration of the Tests

First-year trade apprentices meet at their technical institutes for course tuition in a variety of ways. Generally, groups of 12 - 14 apprentices undertake a block course of two to three weeks, during which they attend the institute full-time. The block courses are repeated regularly

so that all enrolled apprentices may attend. Other trades run a day course every fortnight throughout the year, while some may provide night classes only.

In order to cover the whole range of first-year apprentices in each trade area attending each technical institute, the testing was carried out over a period of several months. In each institute, a senior staff member co-ordinated the testing and was responsible for the administration of the tests to the appropriate groups. Detailed directions for the administration and return of the tests were supplied by the researchers so that testing conditions were as standard as possible for all groups (see Appendix IV).

2.7 Analysis of the Data

All completed test booklets were returned to NZCER for marking and analysis.

(a) Cloze Sections

The cloze passages were scored for exact replacement of the deleted word, with some allowances made for spelling errors. Checks were made of (i) the equivalence in ability of the random subgroups of apprentices to whom the tests were administered, and (ii) the parallelism in difficulty of the test booklets. Cloze test performance was examined in terms of (i) the relative level and spread of test scores for each trade group of 'young adults' (an MMS computer programme adapted from Shoemaker (1973) was employed for this), and (ii) the distribution of apprentices in each trade group attaining each criterion level of 'information gain'.

For the common cloze passages in each booklet (one general passage and one trade passage) additional error analyses of apprentices' responses were carried out to investigate in some detail the nature of the errors made (see Appendix V).

(b) Essay Section

Each essay was assessed independently by two markers. As the writing exercise was brief (10 to 15 lines), an uncomplicated marking scale seemed most appropriate; thus criteria for assessing the essays were based on a scale of 1 to 5 (see Appendix VI.1 for the marking criteria).

The final mark assigned to each essay was the average of the scores awarded by the two markers. Checks on inter-marker agreement and the consistency with which each marker adhered to the marking criteria were made after the marking was completed (see Appendix IV.2 on marker agreement).

3.0 RESULTS

3.1 Sample Characteristics

(a) Distribution by Sex

Table 3.1 summarises the distribution by sex, of secondary school students and apprentices in each of the five trade areas. Apart from Hairdressing, where 91 percent were females, all or virtually all those tested in each trade were male. For secondary school students the numbers of males and females were balanced. The samples reflect the national population balance in each group quite faithfully.

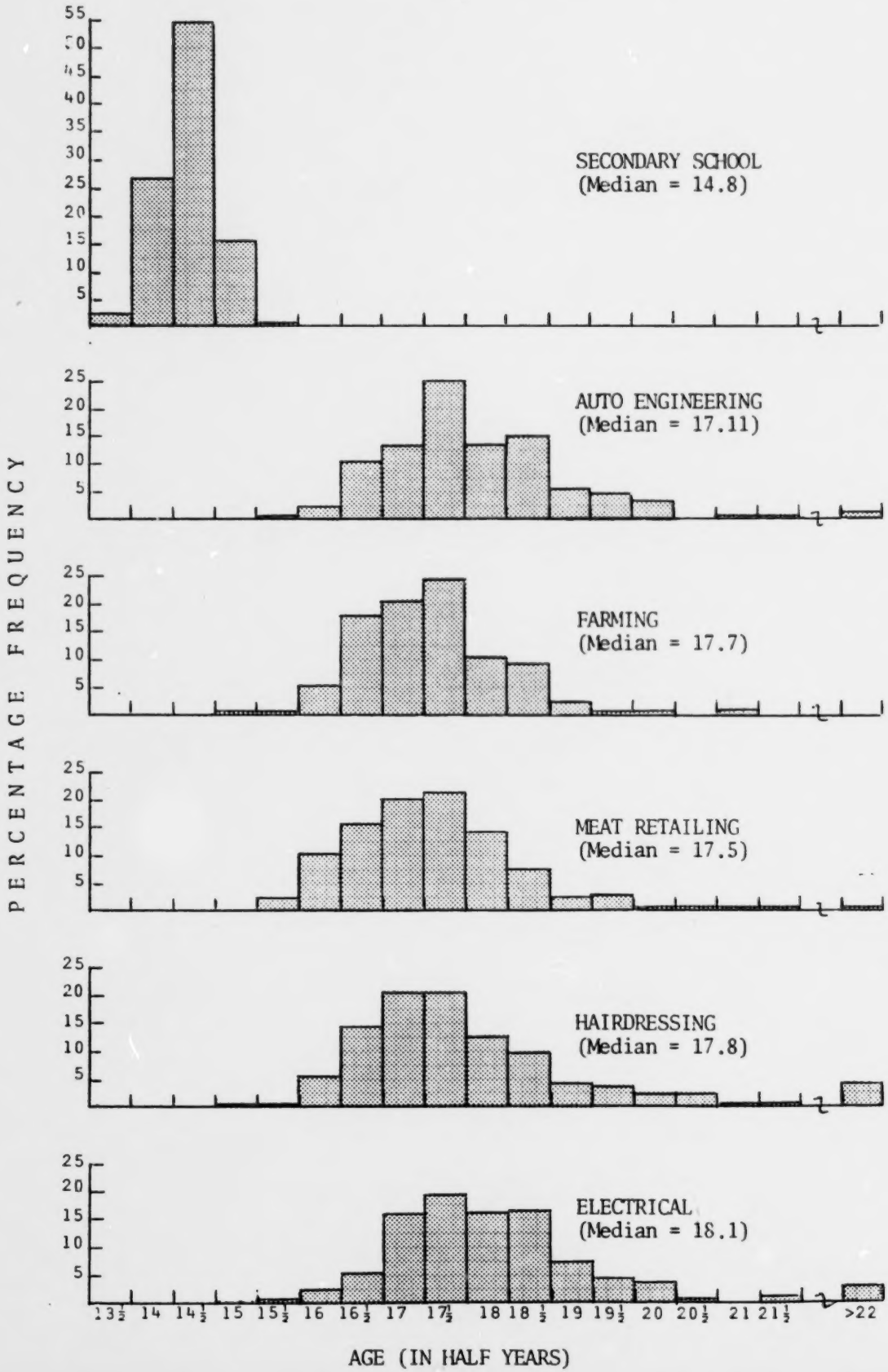
TABLE 3.1 Distribution of the Sample Groups, by Sex

Group	Males		Females		Total
	N	%	N	%	
Secondary Schools	182	*50*	179	*50*	361
Electrical	241	*100*	-	-	241
Hairdressing	30	*9*	291	*91*	321
Meat Retailing	209	*100*	-	-	209
Farming	375	*98*	9	*2*	384
Auto Engineering	332	*99*	2	*1*	334

(b) Distribution by Age

Figure 3.1 reports the distribution of apprentices and secondary school students by age (in half-yearly intervals). The median ages for the apprentice groups were reasonably similar (between 17 and 18 years). As expected, the secondary school sample was younger with a median age of 14.8 years. Each trade area contained a few apprentices over 21 years old.

FIGURE 3.1.: *Distribution of Apprentices and Students by Age.*



(c) Educational Qualifications

Apprentices were asked to provide information on the highest level of educational qualification they had attained. The pattern of educational qualifications differed for each trade area, as would be expected, because of their different entry requirements. Only 6 percent of the sample of electrical apprentices, for instance, were without formal qualifications i.e., they reported having only 2 or 3 years of secondary school attendance, while 60 percent and 45 percent of Meat Retailing and Hairdressing apprentices, respectively, had no formal qualifications. The corresponding figures for Farming and Auto Engineering were 31 percent and 24 percent.

The percentage of apprentices in each trade area attaining qualifications higher than School Certificate was: Hairdressing, 17 percent; Meat Retailing, 17 percent; Auto Engineering, 31 percent: Farming, 39 percent; and Electrical, 64 percent.

The Electrical apprentices held substantially higher educational qualifications than any other trade group (University Entrance, 30 percent; 6th Form Certificate, 25 percent; and School Certificate, 31 percent). This reflects the entry requirement of Electrical trainees (who must be registered in order to practise as Electricians). A summary is provided in Figure 3.2.

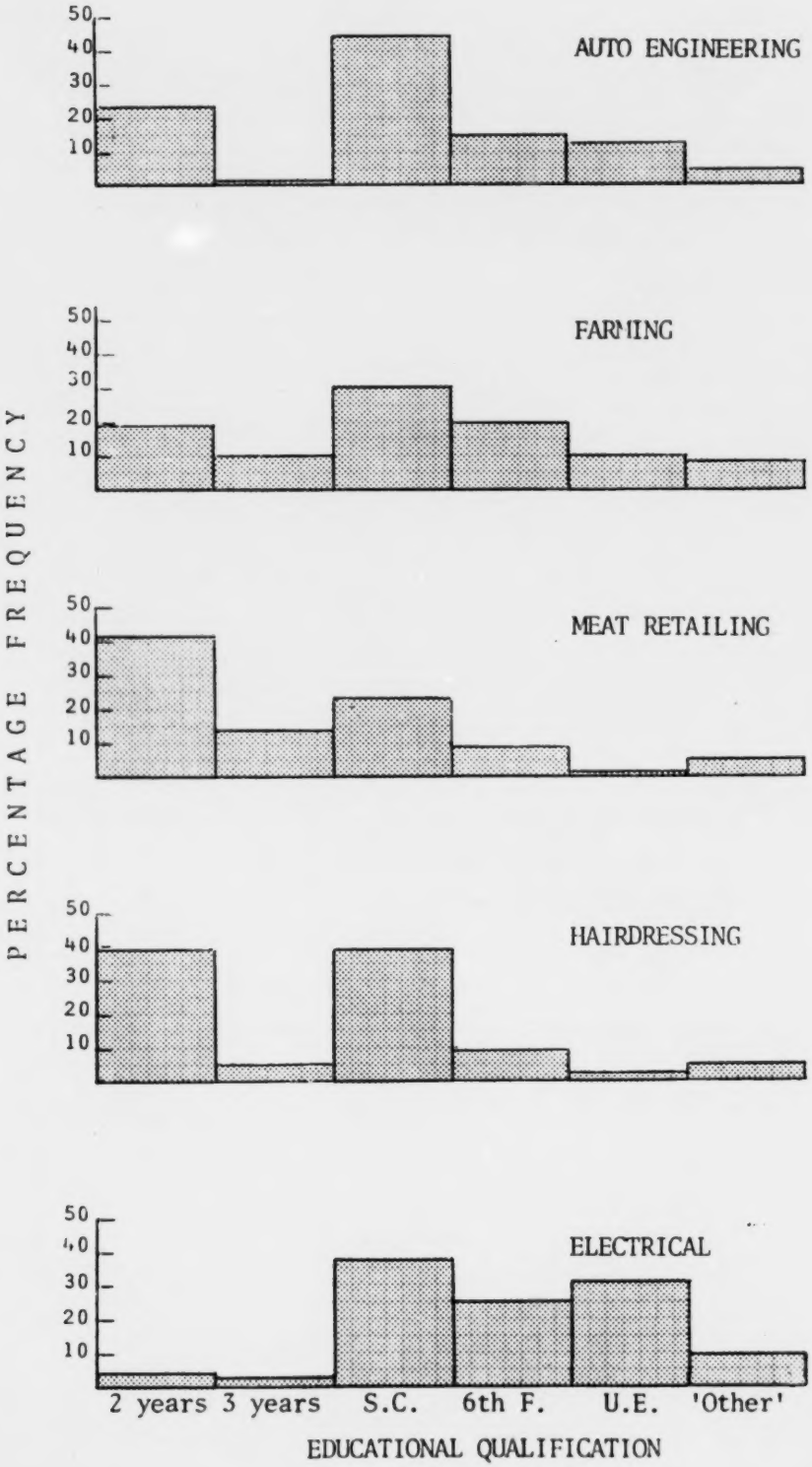
(d) Study Through Correspondence

The apprentices were further asked to report whether they undertook their trade studies through correspondence (i.e., were enrolled with the Technical Correspondence Institute) or attended classes at a technical institute. The percentage of apprentices attending block courses in each trade area and receiving their instructions via correspondence is summarised in Table 3.2.

TABLE 3.2 Number and Percentage of Apprentices Undertaking Study Through Correspondence

Trade Group	Technical Institutes		TCI		Total
	N	%	N	%	N
Electrical	189	*78*	53	*22*	241
Hairdressing	304	*95*	17	*5*	321
Meat Retailing	209	*100*	-	-	209
Farming	322	*84*	62	*16*	384
Auto Engineering	272	*81*	62	*19*	334

FIGURE 3.2.: *Highest Educational Qualification Attained by Apprentices in Each Trade Area.*



The table shows that a least four out of five apprentices attended courses at a technical institute; instruction by correspondence is most popular in electrical, farming and auto engineering trades.

3.2 Performance on the Reading Comprehension Tests [5]

Measures of the apprentices' ability to gain information from reading material were obtained for:

(1) material that young adults might be expected to encounter in their daily living ('general' reading comprehension); and

(2) material that each apprentice group was expected to read during the course of study and work ('trade' reading comprehension).

An important assumption in this study is that the test passages in the 'general' and 'trade' reading comprehension sections represent a balanced sample of the reading materials that these groups of young adults might be expected to encounter in their general day-to-day living, and in their trade-related study and work, respectively.

In this section the performance of each apprentice group on the general and trade reading sections will be considered separately. Performance on the cloze test has been examined in two ways. Firstly, the test scores attained by each group were examined in relation to criterion cut-off scores, which indicated how many apprentices were reading with independence, how many required instruction and how many were experiencing levels of frustration in their reading. Secondly, MMS pooled estimates of the means and standard deviations of scores attained by each apprentice group were computed. These parameters give estimates of how the apprentice group would have performed if all apprentices had answered all 13 test passages in the general section and all seven test passages in the trade section, instead of the four passages actually answered in each section. The reliability of the estimates obtained was consistently very high (greater than 0.94), so considerable confidence may be placed in the precision of the parameters.

[5] Appendices VII and VIII contain comments on the reliability of the test instruments, and the equivalence of the MMS groups and test booklets).

3.3 Reading Comprehension Performance Based on Criterion Scores

One way of interpreting test scores on a cloze test of reading comprehension is to examine performance in relation to established criterion cut-off scores. These scores describe a person's ability to read given material according to three different levels of capability:

(a) with independence, i.e. a cloze score greater than 50 percent;
(b) with instruction or assistance, i.e. a cloze score between 35 and 50 percent; and
(c) with frustration, i.e. a cloze score of less than 35 percent.

(Table 3.3. presents the percentage of apprentices attaining each criterion score in the general and trade reading comprehension sections. This information is also conveyed graphically in Figure 3.3, and the distributions of cloze test scores are represented in Figures 3.4 to 3.9.)

TABLE 3.3 *Percentage of Apprentices in Each Trade Falling Into Each Criterion Cloze Test Score Range*

Trade		Little or No Information Gain (<35%)	Modest Information Gain (35-50%)	Good Information Gain (>50%)
Electrical	General	3.6	20.0	76.4
	Trade	13.1	32.6	54.3
Hairdressing	General	13.0	32.0	55.0
	Trade	8.8	25.9	65.3
Meat Retailing	General	17.0	36.9	46.0
	Trade	15.8	45.6	38.6
Farming	General	12.5	33.3	54.2
	Trade	18.8	58.3	22.9
Auto Engineering	General	13.6	25.4	61.0
	Trade	14.0	31.1	54.9
Secondary Schools	General	12.0	29.6	58.4

FIGURE 3.3.: *Percentage of Apprentices in Each Trade Group Falling Into Each Cloze Criterion Category.*

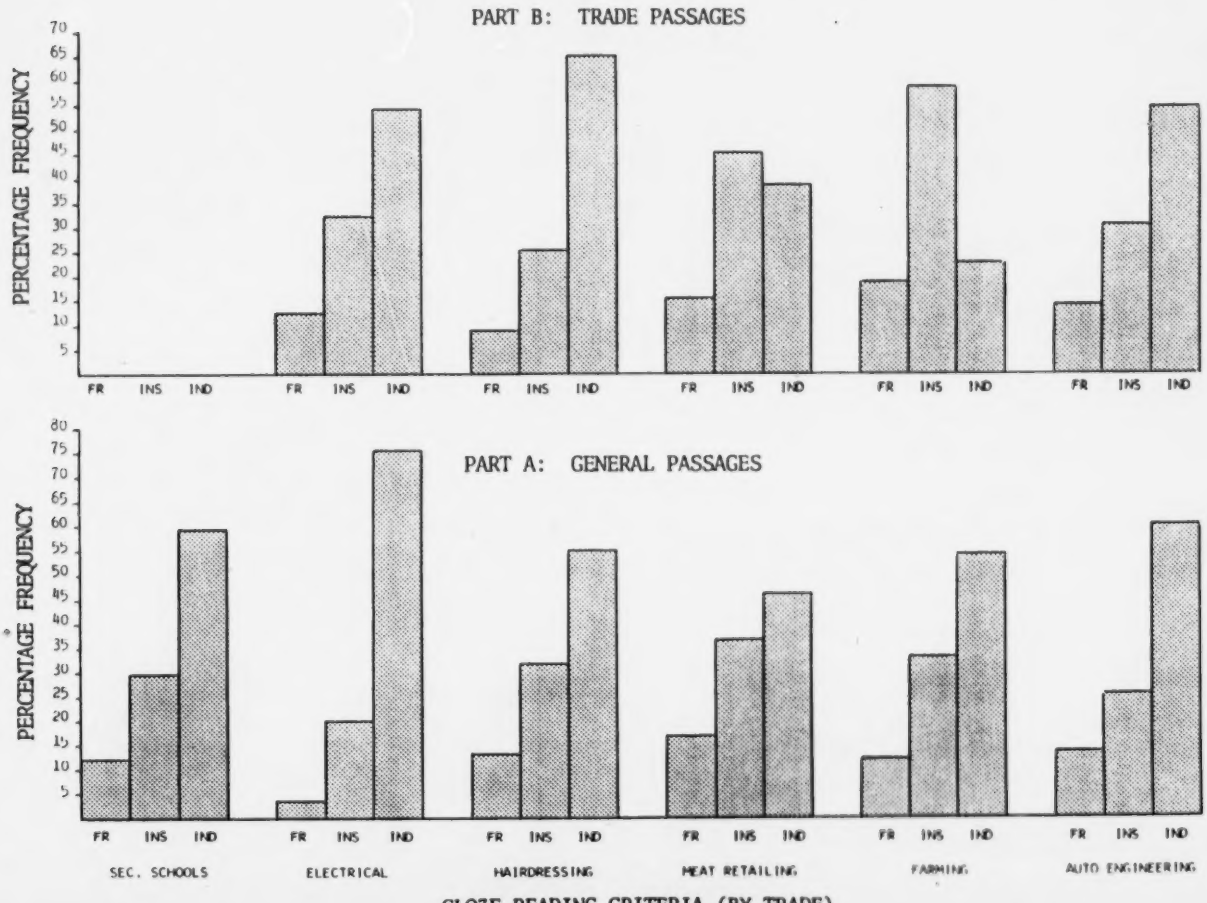


FIGURE 3.4.: *Cumulative Percentage Frequency Distribution of Trade Apprentices on General Passages.*

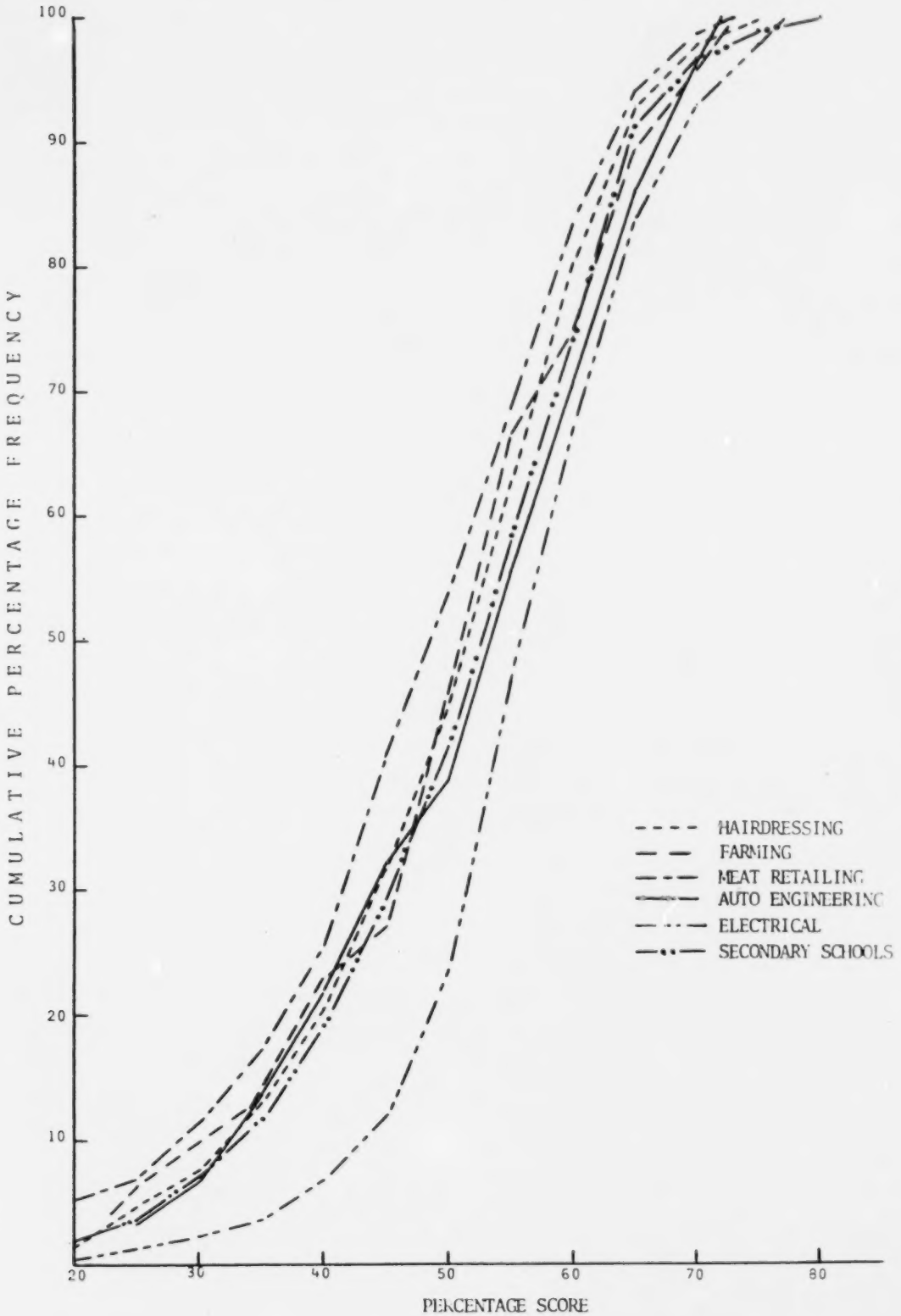


FIGURE 3.5.: *Cumulative Percentage Frequency Distribution of Electrical Apprentices on General and Trade Passages.*

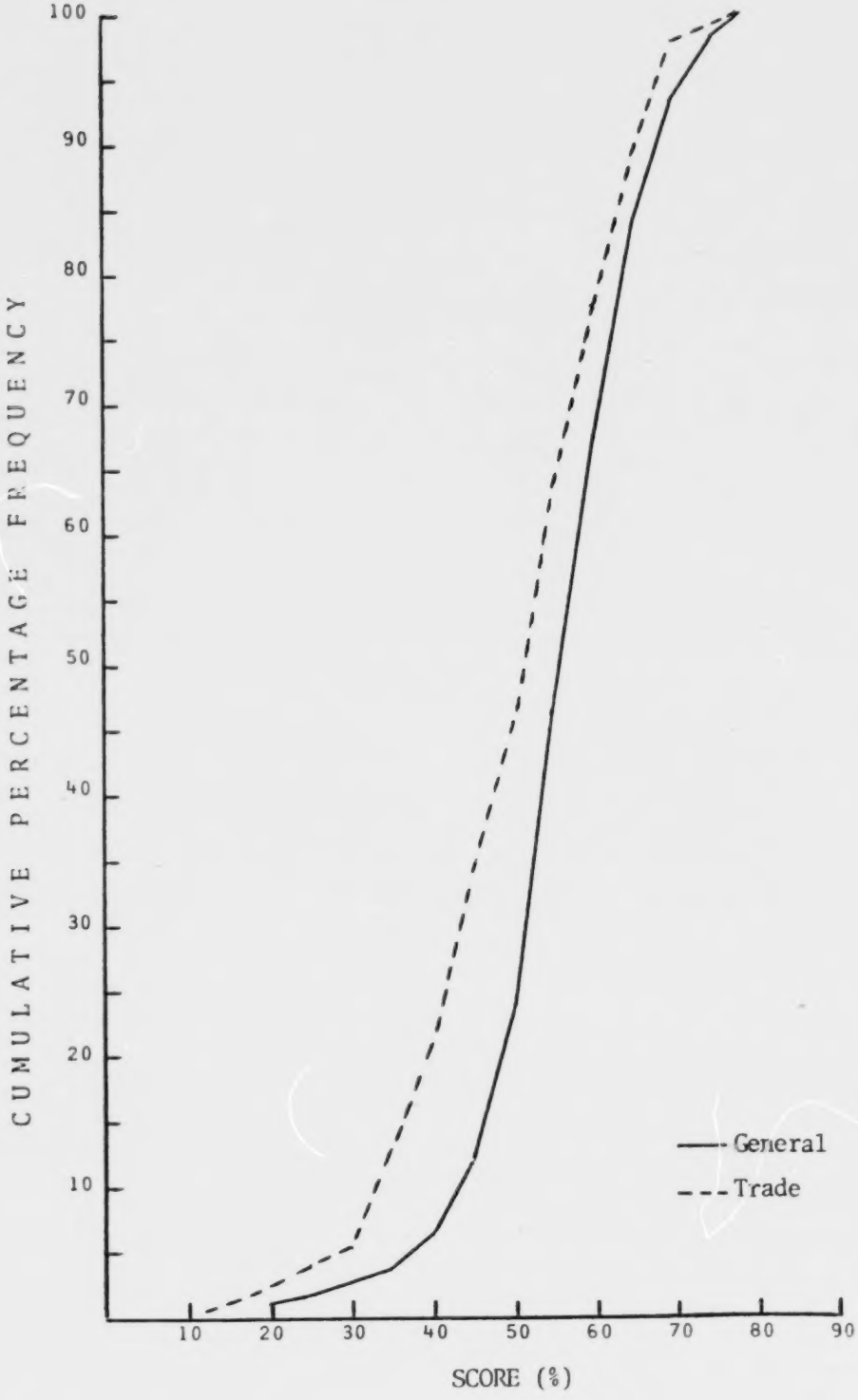


FIGURE 3.6.: *Cumulative Percentage Frequency Distribution of Hairdressing Apprentices on General and Trade Passages.*

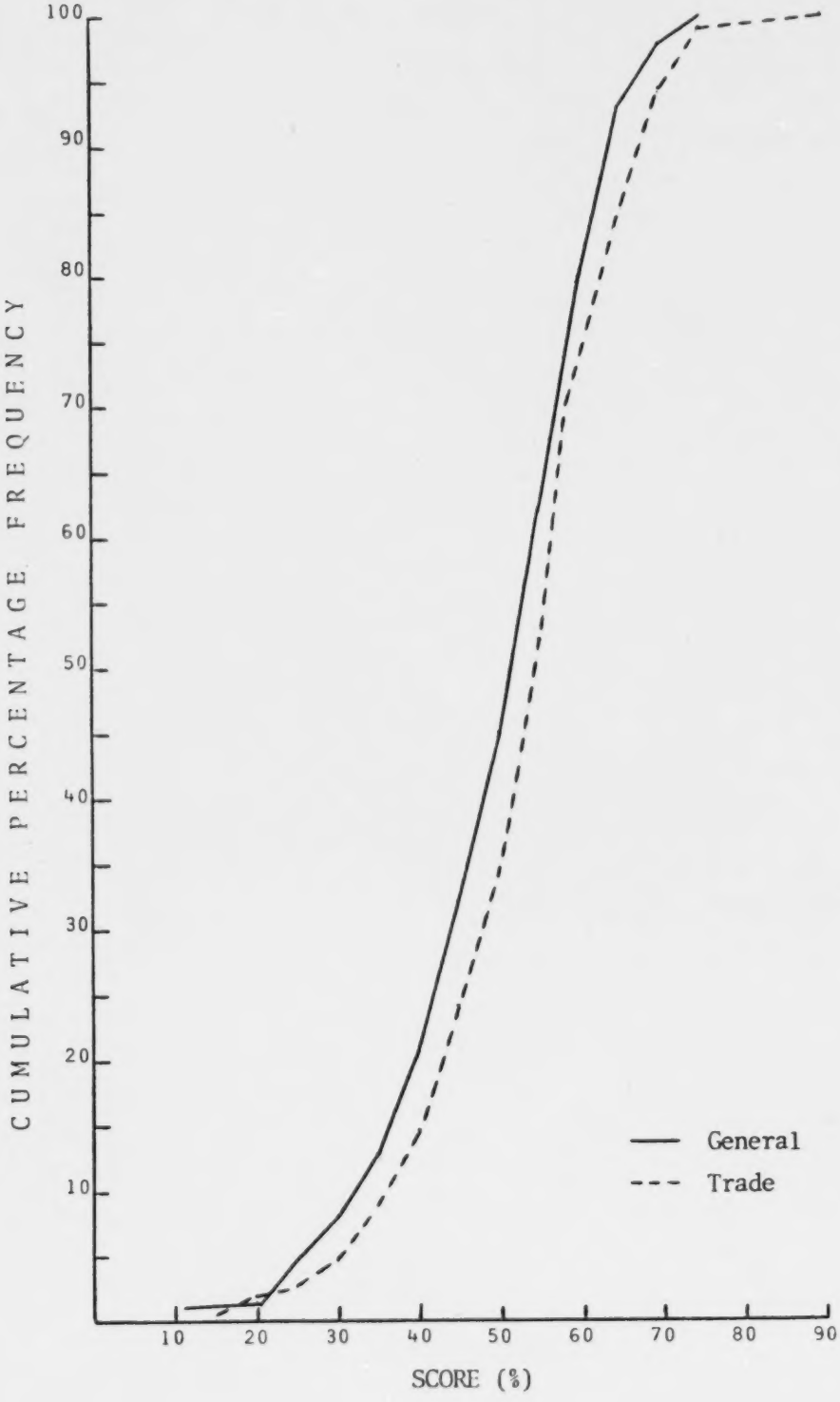


FIGURE 3.7.: *Cumulative Percentage Frequency Distribution of Meat Retailing Apprentices on General and Trade Passages.*

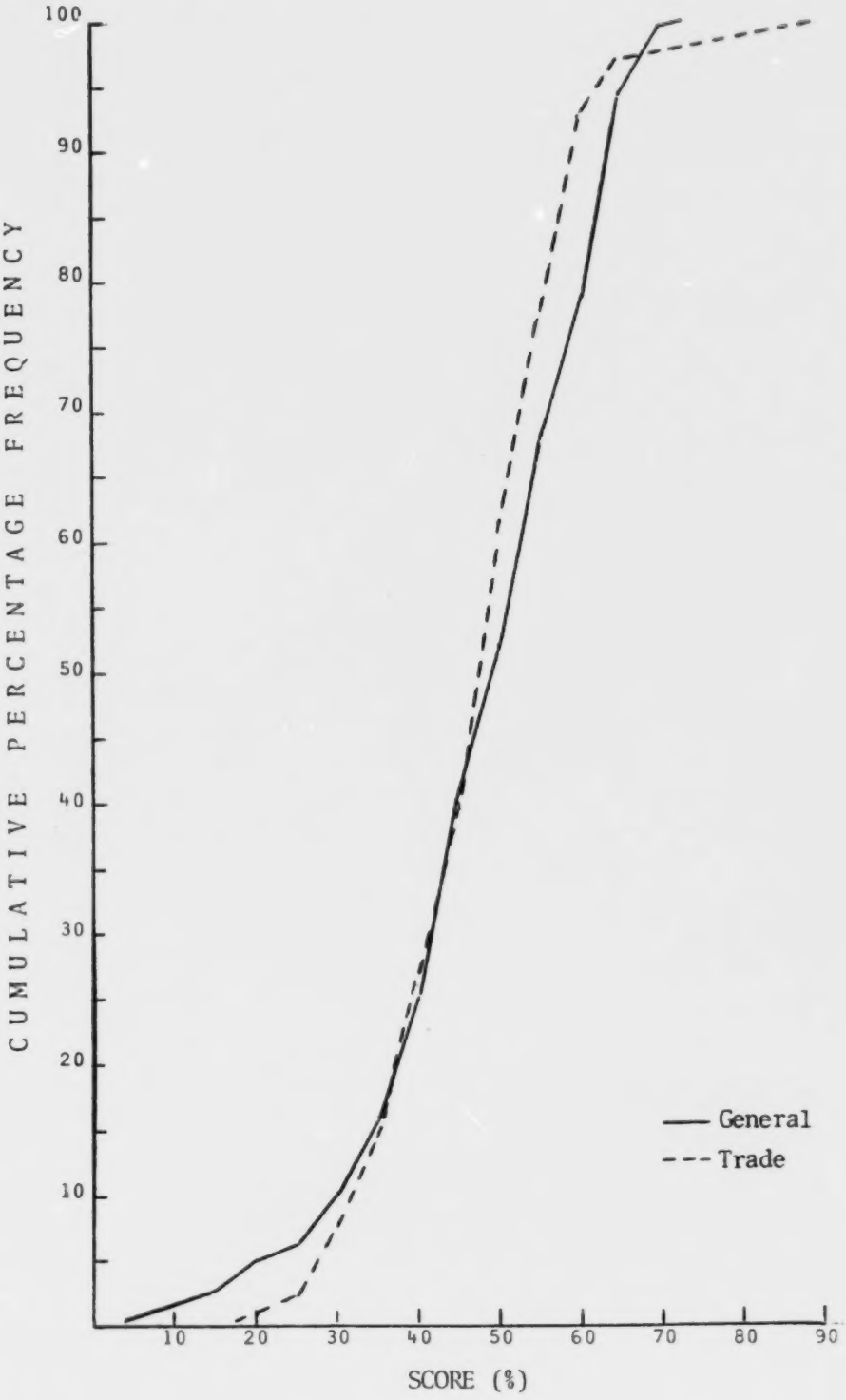


FIGURE 3.8.: *Cumulative Percentage Frequency Distribution of Farming Apprentices on General and Trade Passages.*

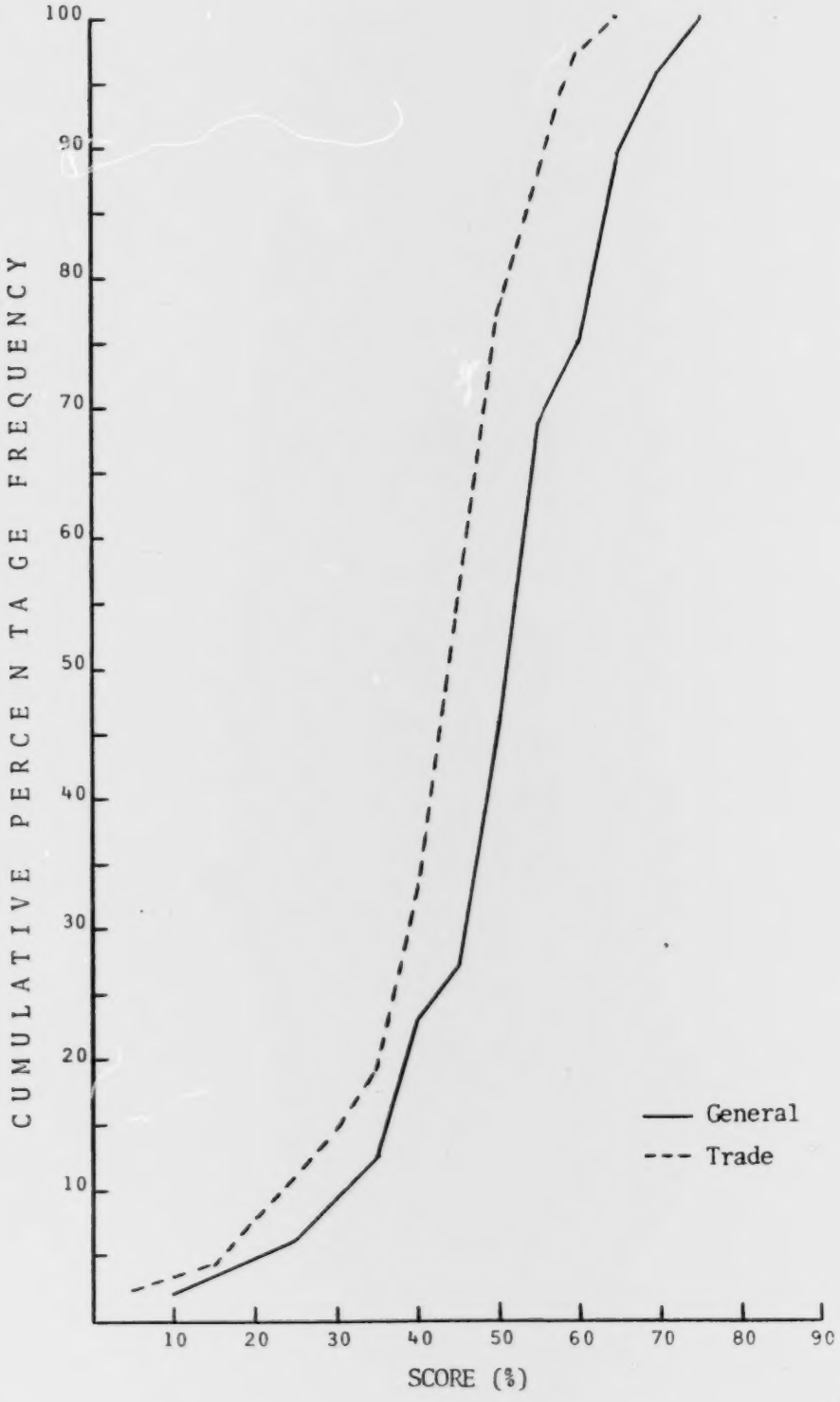
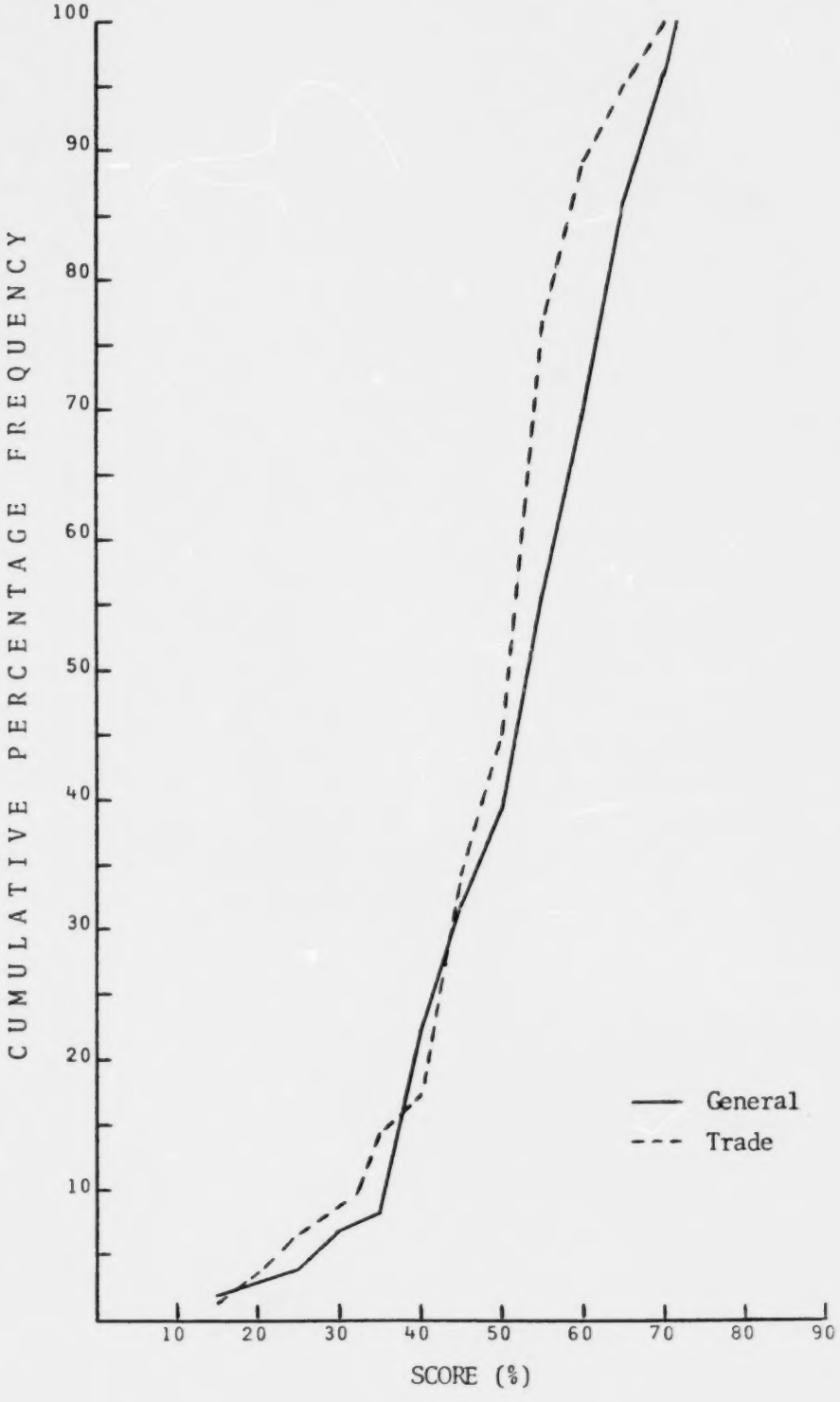


FIGURE 3.9.: *Cumulative Percentage Frequency Distribution of Auto Engineering Apprentices on General and Trade Passages.*



On the general passages approximately 17 percent of the Meat Retailing apprentices obtained a cloze test score of 35 percent or less and fell into the 'frustrated' category, that is, the material was too difficult for them to elicit much meaning from it. Corresponding figures for Hairdressing (13 percent), Farming (13 percent), Auto Engineering (14 percent) and secondary school students (12 percent) showed strong similarities, at a slightly lower level. Very few Electrical apprentices (4 percent) were likely to be frustrated in reading the general passages.

At the other end of the scale, as might be expected, the group containing the largest proportion of able, that is, 'independent' readers demonstrating a good information gain, was the Electrical group (76 percent). They were followed by Auto Engineering apprentices (61 percent) and the group of secondary school students (58 percent). The proportion of the Meat Retailing group reading the general material with ease was substantially lower than all others (46 percent).

Numbers of apprentices falling into each readability category on the trade cloze passages showed a somewhat different pattern. Between 9 percent (Hairdressing) and 19 percent (Farming) of the trade groups fell into the 'frustration' category, likely to experience considerable difficulty with their course study materials. A substantial proportion, from 26 percent (Hairdressing) to 58 percent (Farming), would have been able to read their course materials provided some instructional assistance was provided. From 23 percent (Farming) to 65 percent (Hairdressing) were able to read their course material without too much difficulty.

Overall, for four of the five trades examined (Hairdressing being the exception) the ability of apprentices to read extracts from their respective trade materials was lower than their ability to tackle general reading materials. The greatest disparities seemed to have occurred in Meat Retailing and particularly in Farming.

3.4 Reading Comprehension Scores Based on Means and Standard Deviations

Another, more statistical way of treating the data is to present means and standard deviations, an approach allowing tests of significance to be made between groups. The MMS pooled estimates of means and standard deviations (reported as percentages) for each apprentice group on the general reading comprehension section are reported in Table 3.4.

TABLE 3.4 *Estimated Mean and Standard Deviation of Scores* [1] *on the General Reading Comprehension Section by Apprentices in Each Trade Area.*

Trade	N	Mean	Std. Dev.	Reliability*
Secondary Schools	362	50.6	12.5	.97
Electrical	241	54.4	11.5	.96
Hairdressing	321	48.3	12.5	.97
Meat Retailing	210	43.7	12.2	.97
Farming	384	45.5	11.1	.98
Auto Engineering	334	48.1	13.2	.97

1 Reported as percentages

* Reliability (or precision) of the pooled estimate

An analysis of variance was undertaken to test the statistical significance of the difference in mean scores for each group, and a highly significant F ratio was found ($p < .001$). A further exploration of these differences revealed that the Electrical apprentices performed significantly better than all other trade groups ($p < .01$, using Tukey post hoc comparisons) but only slightly better than the group of secondary school students. Meat Retailing apprentices and Farming cadets performed least well - significantly poorer than Hairdressing and Auto Engineering groups, who attained very similar scores.

This pattern of performance might have been anticipated given the distribution of educational qualifications held by each trade group (see Figure 3.2). However, while the Farming cadets generally held higher educational qualifications than Hairdressing apprentices, the latter performed better.

The estimated mean and standard deviation of performance on the trade section for each apprentice group is summarized in Table 3.5. In isolation, comparative test performance has little meaning. As the content of trade sections is different for each apprentice group, no comparisons between

groups is legitimate. However, if we assume that the passages included in each trade section represent a balanced sample of the materials each apprentice group is expected to be able to read during the course of their study and work, it would appear that the Hairdressing (and to a lesser extent, the Electrical) apprentices were better able to read their course material than the other three groups.

TABLE 3.5 *Estimated Mean and Standard Deviation of Scores[1] on the Trade Reading Comprehension Section by Apprentices in Each Trade Area*

Trade	N	Mean	Std. Dev.	Reliability*
Electrical	241	48.5	12.9	.94
Hairdressing	321	51.2	13.3	.95
Meat Retailing	210	42.6	13.7	.95
Farming	384	40.3	14.1	.95
Auto Engineering	334	44.7	14.3	.95

1 Reported as percentages

* Reliability (or precision) of the pooled estimate

Relationship Between General and Trade Reading Comprehension

The mean test scores obtained by each apprentice group on the general and trade sections were compared to examine the relative performance by each group on the two components. Table 3.6 presents the results of tests of significance between the scores on the two sections for each apprentice group. Different patterns of performance were highlighted for separate trade areas.

TABLE 3.6 *Difference Between the Mean Test Scores[1] for the General and Trade Reading Comprehension Sections for Each Apprentice Group*

Trade	Difference Mean (General) - Mean (Trade)	t
Electrical	5.8	5.26**
Hairdressing	-2.9	-2.79**
Meat Retailing	1.1	0.86
Farming	5.2	5.73**
Auto Engineering	3.4	3.18**

1 Reported as percentages

** clearly significant difference, $p < .01$

For Electrical, Farming and Auto Engineering apprentices, mean performance on the general reading comprehension section was found to be significantly better than on the trade section ($p < .01$). For these groups the typical reading passages selected from their trade course materials were somewhat more difficult than the general reading material encountered in their daily living. Hairdressing apprentices, however, obtained a higher mean score on the trade section than on the general section, indicating greater ease of reading in their course materials. No difference in mean test score was found between the two comprehension sections for the Meat Retailing apprentice group.

In addition to providing valid and reliable measures of an individual's ability to gain information from reading material, cloze test scores are commonly interpreted as measures of the readability or difficulty of textual material. The pattern of mean cloze test scores on the general and trade sections for the five apprentice groups would tend to suggest that the Electrical, Farming and Auto Engineering trade course materials are significantly more difficult to read (have lower readability) than general reading materials, while the reverse is true for Hairdressing course materials.

However, given that the calibre of apprentices differs between trade groups (as evidenced by the differences in mean cloze test scores for the

general section) no direct comparison of the readability of course reading materials can be made between trade groups. A comparison is made possible though, by standardizing the test score distributions for the general section so that all groups have the same mean and standard deviation and are essentially equal in ability. The adjustments applied to each score distribution to achieve this equivalence are then applied to the distribution of cloze scores for the trade section. The resulting adjusted means and standard deviations for the trade reading section are presented in Table 3.7.

TABLE 3.7 *Adjusted Means and Standard Deviations of Scores*[1] *on the Trade Reading Comprehension Section by Apprentices in Each Trade Area*

Trade	Mean	Std. Dev
Electrical	44.2	13.5
Hairdressing	52.9	12.8
Meat Retailing	48.9	13.4
Farming	44.8	15.3
Auto Engineering	46.6	13.1

1 Reported as percentages

Clearly, the Electrical, Farming and Auto Engineering course trade materials have significantly lower readabilities than Meat Retailing materials which, in turn, have significantly lower readability than Hairdressing trade materials ($p < .01$, Tukey post hoc comparisons).

Confirmation of this pattern is provided, in part, by Flesch Readability Indices calculated for each group of reading passages. The resulting 'Reading Ages' for Electrical and Auto Engineering materials are substantially higher (19.5 years and 15.3 years, respectively) than for the remaining three areas of trade materials (Hairdressing, 14.0; Meat Retailing, 13.1; Farming, 13.1).

Additionally, the Flesch Reading Age for the general passages was 15.0 years, suggesting that readability as measured by the structural

features of the materials (e.g., number of syllables per word and average number of words per sentence) on which the index is computed is lower, that is, is more difficult for the trade reading materials for Electrical and Auto Engineering than for the general materials, while the reverse is true of the other trade areas.

3.5 Consistency of Reading Performance in the General and Trade Sections

A measure of the degree of relationship or consistency of the apprentices' reading performances on the general and trade sections of the test is provided by correlational analyses. Table 3.8 summarizes the correlations between test scores on the two sections for each trade group.

TABLE 3.8 *Correlation Coefficients Between the General and Trade Sections*

Test Form	TRADE				
	Electrical	Hairdressing	Meat Retailing	Farming	Auto Engineering
1.1	.70	.75	.57*	.73	.75
1.2	.69	.69	(.25)	.65	.81
2.1	.47	.83	.67	.53	.82
2.2	.65	.83	.82	.68	.34*
3.1	.51	.57	.59	.67	.69
3.2	.42*	.71	.93	.80	.74
4.1	(.15)	.71	.57	.81	.56
4.2	.73	.70	.71	.56	.61
Median	.58	.71	.63	.68	.72

All r's are statistically significant at $p < .01$ or greater except:
( ) not statistically significant
* statistically significant at $p < .05$

All but two of the 40 correlations were significant, suggesting that, in general, there is a high degree of consistency in performance on the two aspects of reading comprehension, general and trade. The median (middle) correlations ranged from 0.58 for Electrical to 0.72 for

Auto Engineering. Those apprentices that scored well on one section generally tended to score well on the other section.

An analysis of the errors made on the common passages in the general and trade reading comprehension sections is contained in Appendix V. The data provide useful information for further study.

3.6 Writing Ability

Each apprentice was required to write 10 to 15 lines on the topic "Why I Became an Apprentice". The scripts were marked for accuracy and clarity of expression rather than for quality of ideas (see Appendix VI.1 for a list of marking criteria).

The mean and standard deviation of essay scores for each trade group are shown in Table 3.9.

TABLE 3.9 *Mean and Standard Deviation of Essay Scores for Each Trade Group*

Trade	Essay		
	Mean	Std. Deviation	n
Electrical	3.3	0.9	202
Hairdressing	3.4	0.7	307
Meat Retailing	3.0	0.7	204
Farming	3.1	0.8	325
Auto Engineering	3.0	0.9	323
Mean (Median)	3.2	(0.8)	

While the mean scores between trade groups were quite similar (ranging from 3.00 to 3.37), an analysis of variance indicated that some of the differences were statistically significant ($F = 6.07$, $p < .001$). The Hairdressing apprentices (mainly females) and the Electrical apprentices (the group with the highest overall level of educational qualifications) demonstrated significantly better writing/communication skills, in terms of language usage, coherence and expression of ideas, and the construction of sentences and paragraphs ($p < .01$ using Tukey post hoc comparisons). There were no appreciable differences in the writing abilities of the Meat Retailing, Auto Engineering and Farming apprentices.

3.7 Relationship between Reading Comprehension and Writing Ability

The relationship between apprentices' reading comprehension and writing abilities was obtained by correlating the scores for the two measures. The median correlations between the scores for writing and for trade and general reading comprehension over the five trade areas are reported in Table 3.10.

TABLE 3.10 *Median[1] Correlation Coefficients Between the Essay and General and Trade Reading Comprehension Sections*

Trade	Reading Comprehension	
	General	Trade
Electrical	0.29	0.26
Hairdressing	0.53	0.52
Meat Retailing	0.40	0.34
Farming	0.62	0.50
Auto Engineering	0.43	0.41

All r's were significant at $p < .05$ or beyond

[1] Median r over the parallel test forms

The median correlations between the essay and the general reading comprehension section were only marginally higher than those between the

essay and the trade cloze test scores and were weakest for the Electrical apprentices. In general, there was a moderate (although statistically significant) level of relationship between the reading and writing abilities of apprentices, suggesting that those apprentices that read well tend to be able to write well also.

3.8 Reading Comprehension, Writing and First Qualifying Examination Performance

The score attained by apprentices in their first qualifying examination taken at the end of 1982 was inspected in relation to their reading comprehension and writing abilities. (Farming cadets were excluded from these analyses as they are not formally assessed in their first year of training.) The median correlation coefficients between the general and trade cloze test scores, essay mark and first qualifying examination performance are summarized in Table 3.11.

TABLE 3.11 *Median[1] Correlation Coefficients Between Reading Comprehension (RC), Essay and First Qualifying Examinations*

Trade	Median Correlation Between Examination And		
	General RC	Trade RC	Essay
Electrical	0.06	0.01	0.15
Hairdressing	0.42	0.45	0.34
Meat Retailing	0.62	0.50	0.39
Auto Engineering	0.39	0.49	0.25

[1]Median r over parallel test forms

For Electrical apprentices, no correlation was found between reading comprehension and examination performance. This lack of relationship may reflect the nature of the examination or may be due to the homogeneous nature of the group of Electrical apprentices. Apprentices who enter the trade course with University Entrance passes in certain subjects are exempted from sitting and the group presenting for the examination therefore is composed of those apprentices with lower entry qualifications.

For Hairdressing, Meat Retailing and Auto Engineering apprentice groups, a reasonably strong relationshipexisted between cloze reading

ability and examination performance. A slightly but consistently weaker correlation was found between the essay and examination scores.

PART II

4.0 METHODOLOGY

4.1 Sample

To undertake a readability analysis of the 1982 First Qualifying TCB examination papers it was clearly not possible to use the 1982 candidates themselves! Comparable samples of apprentices were therefore selected from the 1983 first-year intakes in Ladies Hairdressing, Automotive Engineering, Electrical and Meat Retailing trades at three technical institutes in Auckland and Wellington during the third term of the year. Farming apprentices were not included in this part of the study as they are internally assessed and do not sit an external examination paper.

In selecting the 1983 groups, consideration was given to sample size and representativeness. Analysis of the 1982 results indicated that there were no significant differences between the performances of apprentices in these trades from different technical institutes or community colleges. Therefore, for administrative convenience the 1983 apprentices were selected from Auckland and Wellington technical institutes only. In Meat Retailing the number of apprentices accessible at the time of testing was smaller than expected. A group of Maori apprentices undertaking the same course was therefore included to provide an adequate sample size. The number of apprentices tested in each trade area was: Electrical, 39; Auto Engineering, 32; Ladies Hairdressing, 35; and Meat Retailing, 43.

A check on the distribution of the 1983 samples by sex, age and educational qualifications showed them to be almost identical in composition to the 1982 samples. Their performances on a common 'general' and a common 'trade' passage (also admsinistered to the 1982 groups) were quite comparable. For all intents and purposes then, it can be assumed that the 1982 and 1983 samples were equivalent and representative of their respective trades (see Appendix IX).

4.2 The Test Booklets

For each trade group a separate test booklet was prepared, consisting of the common general and common trade passages used in Part I of the study plus three extracts from the beginning, middle and end of each

1982 trade examination paper. The examination extracts were 'clozed' in the same way as for Part I (See Appendix I.2). (The common passages were included so that a check could be made on the comparability of the 1982 and 1983 samples).

4.3 Administration of the Tests

The testing programme within each institute was co-ordinated by a senior staff member. The tests were administered to the apprentices by a trained research assistant who used written directions so that the administration and test conditions were as standard as possible for all groups.

4.4 Analysis of the Data

All completed test booklets were marked and analysed by NZCER staff. Similar procedures to those used in Part I were used to mark the cloze passages.

Preliminary checks were made on the comparability of the 1982 and 1983 samples (see Appendix IX). Cloze performance of the examination extracts was examined in two respects: (1) the distribution of apprentices in each trade attaining each reading criterion level and (2) the readability of the examination (i.e. means and standard deviation of test scores gained by apprentices in each trade group). In addition, the format of the examination papers was inspected for possible explanations for cloze performance.

5.0 RESULTS

5.1 Examination Cloze Performance based on Criterion Scores[7]

The cloze tests scores were examined in relation to established criterion cut-off scores to determine the proportions of apprentices falling into the Frustrational, Instructional and Independent levels of reading ability. These are reported in Table 5.1 and presented graphically in Figure 5.1.

---

[7] Appendix VII contains details of the reliability of the tests.

TABLE 5.1 *Percentage of Apprentices in Each Trade Group Falling into Each Criterion Cloze Test Score Range.*

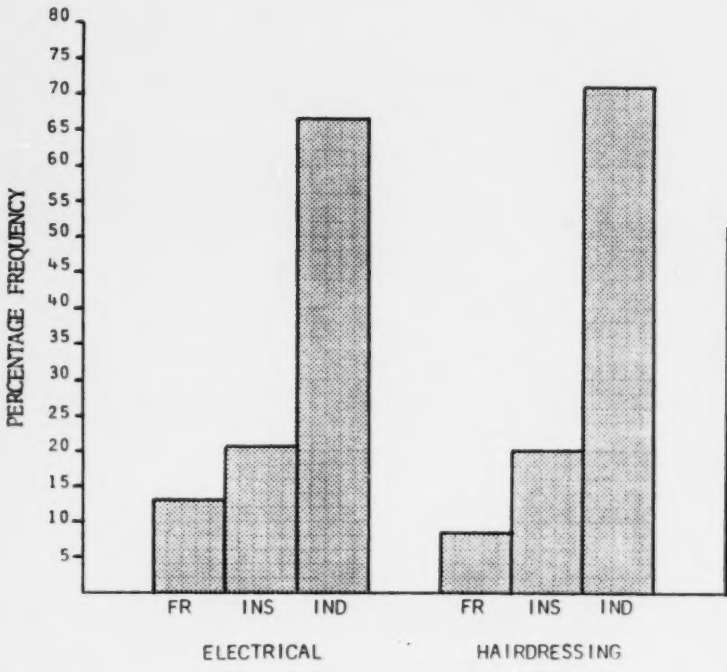
Trade	Little or No Information Gain <35%	Modest Information Gain 35 - 50%	Good Information Gain >50%
Electrical	12.8	20.5	66.7
Hairdressing	8.6	20.0	71.4
Meat Retailing	51.2	30.2	18.6
Auto Engineering	18.8	25.0	56.2

The data show clearly that the Meat Retailing group has considerably more apprentices scoring at the Frustrational level (51 percent) than the other three trades. The position becomes even more serious when the second column (Instructional level) is added to the first column; over 80 percent of Meat Retailing apprentices need some assistance in reading their examination papers. It may be that the low scores are related to the tabulated question format used in the Meat Retailing examination. These results would seem to suggest that most Meat Retailing apprentices required to sit the final examination in 1982 found that it was too difficult to read.

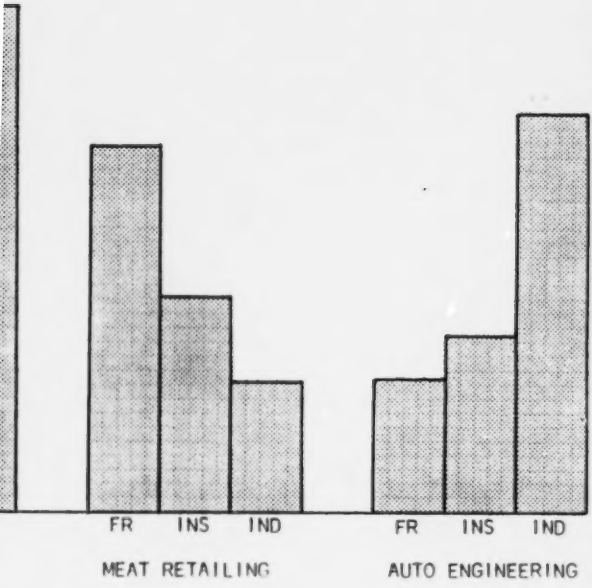
The results would also suggest that a moderately large number of Auto Engineering apprentices cannot read their examination papers independently (i.e. without assistance). If this is so, performance in this examination in 1982 would be as much an indication of ability to understand the examination questions as it was an indication of subject knowledge.

While one might expect some apprentices to score at the frustrational level, the Ladies Hairdressing examination paper, and to a lesser extent the Electrical paper, are close to being acceptable on the assumption that about a 75 percent criterion level for readability is desirable, that is, the papers can be read 'independently' by at least three quarters of the

FIGURE 5.1.: *Percentage of Apprentices in Each Trade Category on Examination Cloze Performan*



h Trade Group falling into each Criterion
rformance.



G CRITERIA (BY TRADE)

candidates.

5.2 Examination Readability based on Means and Standard Deviations

In estimating the readability of a textbook it is usual to take three extracts from the book, assess the readability index for each, find the average or mean score and use this as the book's overall index of readability. In the present study three extracts were taken from the beginning, middle and end of each trade examination paper; in terms of sampling they represent a far higher proportion of the actual text than would be the case in estimating textbook readability. A measure of the readability of each trade examination was obtained from the mean performance of the apprentices on the respective examination extracts. The means and standard deviations of scores on these extracts are reported in Table 5.2. The validity of the cloze as a measure of examination readability is examined in Appendix X.

TABLE 5.2 *Means and Standard Deviations of Scores on the Examination Extracts by Apprentices in Each Trade*

Trade	Mean	Std. Dev.	N
Electrical	31.7	10.3	39
Hairdressing	32.1	8.7	35
Auto Engineering	28.6	11.1	32
Meat Retailing	21.5	9.7	43

Again, direct comparisons between the trade groups are not legitimate because of the differing content in each examination paper. However, as for Part I, the reading ability levels of apprentices in each trade may be statistically adjusted by using the common general reading passage as the anchor. In doing this the assumption is made that the apprentices in each trade have equal reading ability. The relative readability of the examinations can therefore be gauged, independent of the reading abilities of the apprentices currently taking them.

TABLE 5.3 *Adjusted Means and Standard Deviations of Scores on the Examination Extracts by Apprentices in each Trade Area*

Trade	Mean	Std. Dev.
Electrical	29.1	12.1
Hairdressing	30.4	6.6
Auto Engineering	27.8	10.0
Meat Retailing	27.9	7.4

Table 5.3 presents the adjusted means and standard deviations (indices of readability) for each trade group on their examination extracts. When adjusted in this way, the examination papers appear to be very similar in readability, as differences between the means for each paper were statistically non-significant. It would seem that TCB examiners are currently preparing examination papers of relatively uniform levels of readability across the sampled trades, but that in practice, this appears to penalise some groups more than others.

It is important to consider several factors when trying to disentangle the effects of reading ability, content knowledge and features of the examination (such as, question format) on examination performance. Firstly, the "reading load" of examinations (that is, the amount of material that is required to be read by apprentices to answer the examination questions) differs between trades. The extracts selected for inclusion in the cloze test were all of similar length (i.e. of between 160 and 180 words). However, these extracts actually represented different proportions of the trade papers. For instance, the cloze extracts from the Electrical examination represented 47 percent of the paper while the corresponding figures for the Ladies Hairdressing, Auto Engineering and Meat Retailing papers were 87 percent, 50 percent and 54 percent, respectively. Hence the total 'reading load' of the Hairdressing exam was substantially less than that in the other three trade papers. The Electrical apprentices were required to read the greatest amount of material in order to answer their examination questions.

Secondly, the reading ability level of apprentices differs markedly

between trade groups (see Table 3.4, Part I). Electrical apprentices as a group appear to be more able readers than apprentices in the other trade groups sampled. Meat Retailing and Farming apprentices have much more difficulty in reading the same materials. What then are the combined effects of these two factors?

Electrical apprentices demonstrate a reading ability reasonably well in line with the demands of their examination, in terms of reading load. However, Meat Retailing and Auto Engineering apprentices appear to be somewhat disadvantaged by the reading requirements of their examinations. They must read a similar amount of textual material to the Electrical apprentices, but as they tend to be less able readers, in general, they do not cope as well. Hairdressing apprentices, on the other hand, would seem to be in a rather better position by taking an examination with a very light reading load.

Thirdly, other features of the examination, such as, general vocabulary (i.e. non-trade vocabulary), language structure and question format could also be expected to influence examination performance. For example, the format of the questions used differed markedly among the examination papers. This, in effect, not only influences the 'reading load' for the apprentices but also influences their 'writing load'. The Electrical examination required apprentices to write answers to all questions; for the other examinations from 10 to 20 percent of the questions were of a recognition type, i.e. multiple-choice or matching. In addition, only the Electrical examination offered a choice of questions, whilst all other apprentices were required to answer all questions. The Meat Retailing and Ladies Hairdressing papers provided spaces on the paper itself for the answers to be written, but candidates in the two other trades were required to use answer booklets. The Meat Retailing exam appeared to require the least writing from the apprentices, with the majority of questions asking for one or two word answers. However, the highly tabulated question format may have contributed to the low estimates of readability for this exam. The Ladies Hairdressing was similar, although spaces provided for answers tended to be a little larger, implying that longer answers were expected. On the whole the writing requirements for apprentices in Electrical and Auto Engineering were far greater than they were for the other two trades.

Finally, the 'cognitive level' of questions in the Electrical paper differed substantially from that of questions in the papers of the other three trades. For Meat Retailing and Ladies Hairdressing and, to a lesser extent, Auto Engineering, the apprentices were asked simply to recall information in answer to examination questions, but Electrical apprentices were required to demonstrate higher level cognitive skills, such as interpreting information and applying it to given situations. Moreover, far more computations and diagrammatic representations of information were demanded by the examiner.

It is very difficult to determine whether apprentices failed to fully understand their examination papers because of their lack of general reading ability or specific subject-matter content knowledge or their inability to cope adequately with specific types of questions. The interaction of these three factors, plus others, such as, motivation, has contributed to the readability of the exam papers. Evidence from results presented in earlier sections of this report would tend to suggest that poor comprehension of the exam papers is due in part to poor reading ability generally.

CONCLUSIONS AND RECOMMENDATIONS

6.0 CONCLUSIONS

6.1 Comprehension of General and Trade Reading Materials

In order to make generalizable comments about the ability of first year apprentices in the five trade areas to understand (i.e. gain information from) general and trade reading materials two assumptions were made. Firstly, it was assumed that the reading passages in the cloze tests represented a balanced sample of the reading materials that apprentices may be expected to encounter in their daily living and in their respective trade areas. Secondly, it was assumed that the apprentices participating in the study were representative of the population of first-year apprentices in each trade. Given the careful procedures adopted for selecting both reading materials and apprentice groups, it is highly likely that these conditions have been satisfied.

The performance of apprentices on the tests of reading comprehension may be summarized as follows:

1. Apart from the Electrical apprentices, the apprentices in four of the trade groups (Meat Retailing, Hairdressing, Farming and Auto Engineering) had a lower overall level of reading ability than the group representing the 'general population', that is, Form 4 and 5 secondary school students. This was not unexpected given the educational backgrounds and known past performance of these groups.

2. For four of the five trades examined (excluding Hairdressing) the ability of apprentices to read extracts from their respective trade materials was much lower than their ability to read general materials.

   While it is difficult to disentangle the relative readability of the materials in the general and trade areas because of their dissimilarity in writing style, vocabulary and format, it would seem that, in general, trade materials were somewhat more difficult to read than general reading materials. It is somewhat unrealistic to consider the readability of course materials in isolation from the calibre of apprentices who enrol for study in each particular trade. Therefore, a statistical adjustment was made for differences in the reading ability of apprentice groups. It was found that the Electrical, Farming and Auto Engineering trade materials were substantially more difficult than Meat Retailing and Hairdressing trade materials.

3. About 15 percent of the apprentices in each trade (9 percent in Hairdressing) were experiencing considerable difficulty with their course study materials, that is, they gained little or no information from them. A substantial proportion, from 26 percent (Hairdressing) to 58 percent (Farming), were able to read their course materials provided some instructional assistance was available. From 23 percent (Farming) to 65 percent (Hairdressing) were able to read their course materials with few problems.

   The trades which demonstrated the greatest disparity between the readability of training course material and the reading ability of the apprentices enrolled would seem to be Farming and Meat Retailing.

6.2 Relationship Between Reading Comprehension, Writing and Examination Performance

The Hairdressing and Electrical apprentices demonstrated somewhat better writing/communication skills than the remaining three trade groups. This finding might have been anticipated considering the higher educational qualifications held by the Electrical apprentices and the nature of the Hairdressing group (that, is the group was largely female). Research has generally found that females write better than males.

In addition, those apprentices who read well tended to write well also. The relationship between reading and writing was somewhat stronger for Hairdressing and Farming apprentices than for the other groups.

Apprentices' ability to read tended to influence (correlate with) end-of-year examination performance to a greater extent than their ability to write. The Electrical apprentices were the exception to this finding with no correlation existing between either reading or writing and end-of-year exam performance. This may have been due to the more selective nature of the Electrical apprentices, and to the greater proportion of exam questions which required computational and diagrammatic answers, as opposed to extended written answers.

6.3 Examination Readability

The results seem to suggest that TCB examiners in Electrical, Automotive Engineering, Hairdressing and Meat Retailing have prepared examination papers of relatively uniform difficulty in terms of readability. However, when considering the different levels of reading ability amongst the four trades, this uniformity has tended to penalize some groups more than others.

Meat Retailing apprentices had substantially more difficulty in comprehending their examination than the other three trades. Over 80% of the apprentices appeared to need some assistance in reading their examination paper. The highly tabulated question format may have contributed heavily to this performance.

A moderately large number of Automotive Engineering students also could not read their exam without assistance. For these two groups, performance in the 1982 examinations was as much an indication of ability to understand the exam questions as an indication of subject knowledge.

The Electrical and Hairdressing examination papers were closer to an acceptable readability level with almost 75% of the groups being able to read the examination independently.

Reasons for the different levels of readability in the four trade papers are not straightforward to establish.

Examination scores are influenced by the interaction of a number of factors such as, level of content knowledge, the 'reading and writing loads' of the paper, the 'cognitive level' of questions and the format of questions as well as apprentices' reading and writing abilities. In general, the "demands" of the examinations in terms of these factors were commensurate with the demonstrated reading and writing abilities of the apprentices in each trade. Underlying the low readability of the examination papers would appear to be general poor reading ability.

7.0 RECOMMENDATIONS

On the basis of the findings of this study a number of suggestions for future practice in setting TCB examinations may be put forward.

It is recommended that:

1. Trade tutors be made aware of the results of this study.
2. Other trades not covered by this report also be investigated for the match or mismatch between apprentices' reading abilities and their trade study materials and national examinations.
3. Tutors seeking to find this match/mismatch between apprentices' reading ability and their study materials be encouraged to use a readability index of known validity, reliability and ease of use, such as the cloze procedure as described and used in this report or a predictive index,

such as, the Mugford Charts or the Forcast formulas.

4. National TCB examiners be advised to take into consideration the relative reading abilities of apprentices when setting end-of-year national examinations and be encouraged to use an appropriate index to estimate their readability.

5. National TCB examiners bear in mind the direct relationship between apprentices' reading and writing abilities when selecting the types of objective or subjective questions for their papers.

6. Assistance be given to TCB examiners in preparing valid national examination papers in the light of the likely reading and writing abilities of apprentices.

7. The role of the moderator or moderators of TCB national examinations be extended and clearly specified. In addition to scrutinizing the examination for validity in terms of content coverage, the moderator(s) should also scrutinize the exam papers for the technical qualities which influence examination difficulty, such as, clarity of directions, readability, reading and writing loads, appropriateness of question type etc. For moderating the technical qualities of an examination paper, the moderator does not need to be a subject specialist but needs to have expertise in educational assessment.

   The examiner and moderator(s) should be jointly accountable to TCB and exam candidates for the quality of the examination paper.

REFERENCES

Anderson, J. (1969) 'Cloze Procedure as a Measure of Readability'. Papua and New Guinea Journal of Education. 6, (3), 21-8.

Boag, P. (1975) 'Reading in the Secondary School'. Education, 24, 1.

Bormuth, J.R. (1962) Cloze Tests as Measures of Readability. Unpublished Doctoral Dissertation, Indiana University.

Bormuth, J.R. (1965) "Optimum Sample Size and Cloze Test Length in Readability Measurement", Journal of Educational Measurement, 2, 111-116.

Bormuth, J.R. (1966) 'Readability: A new approach', Reading Research Quarterly, 1 (3), 79-131.

Bormuth, J.R. (1967) Implications and Use of the Cloze Procedure in the Evaluation of Instructional Programs. Occasional Report No. 3, Center for the Study of Evaluation of Instructional Programs, University of California: Los Angeles.

Bormuth, J.R. (1968) Cloze Test readability: criterion reference scores. Journal of Educational Measurement. 5, 189-96.

Bormuth, J.R. (1969) "Factor Validity of Cloze Tests as a Measure of Reading Comprehension Ability". Reading Research Quarterly, 4, Spring, 358-365.

Bormuth, J.R. (1970) On the Theory of Achievement Test Items. Chicago, Ill.: The University of Chicago Press.

Bormuth, J.R. (1971) Development of Standards of Readability: Toward a Rational Criterion of Passage Performance. Final Report. Project No. 9-0237. Bureau of Research, U.S. Office of Education.

Bormuth, J.R. (1973) 'Reading Literacy: its definition and assessment', Reading Research, 9 (1), 7-66.

Bormuth, J.R. (1975) 'The Cloze Procedure-Literacy in the Classroom' in Page, W.D. (Ed). Help for the Reading Teacher, New Directions in Research, ERIC Clearinghouse on Reading and Communication Skills, Illinois.

Callaghan, C. (1981) 'Developing Distance Education'. Tutor, 23.

Elley, W.B. (1975) Assessing the difficulty of Reading Materials: The Noun Frequency Method. Wellington: New Zealand Council for Educational Research.

Elley, W.B. (1976) 'The "Cloze" Procedure - A method of Testing, Teaching and Assessing Readability', in Doake, D.B. and O'Rourke, B.T. New Directions for Reading Teaching, International Reading Association.

Elley, W.B. (1978) "Review of the Cloze Procedure" in Buros (Ed). The Eighth Mental Measurements Yearbook, 11, 74-6.

Elley, W.B. and Marriot, R. (1983) The Teaching and Testing of Trade Trainees, paper presented at the NZARE conference, Wellington.

Elley, W.B. and Marriot, R. (1984) Apprentices Learning in Block Courses: A research report, Department of Languages and Communication Studies, Christchurch Polytechnic.

Forbes, F.W. and Cottle, W.C. (1953) 'A New Method for Determining Readability of Standardized Tests', Journal of Applied Psychology, 37, 185-190.

Hansen, L.H. and Hesse, K.D. (1974) A Pilot Reading Literacy Assessment of Madison Public School Students: final report, Dept. of Research and Development, Madison, Wisconsin Public Schools.

Harrison, C. (1980) Readability in the Classroom, Cambridge University Press, Cambridge.

Haynes, E.F. (1978) 'Using Research in Preparing to Teach Writing'. English Journal, 67, 82-88.

Holmes, N. (1981) 'The Readability of Study Materials: Recent Research in New Zealand' in Daniel, J.S., Stroud, M.A. and Thompson, J.R. (Eds). Learning at a Distance: a world perspective. Proceedings of the International Council for Correspondence Education, Athabasca University, Edmonton, Canada, 109-113.

Karnes, F.A. and Gunn, C.N. (1976) 'Analysis of Reading Teachers Perceptions of the Reading Programs in Post-secondary Vocational-Technical Complexes in Mississippi'. Vocational Reading Development Program No. 34, Hattiesburg, Mississippi: Bureau of Educational Research, University of Southern Mississippi.

Lord, F.M. (1962) 'Estimating Norms by Item Sampling'. Educational and Psychological Measurement, 22, 259-267.

MacGinitie, W.H. (1961) 'Contextual constraint in English prose paragraphs'. Journal of Psychology, 51, 121-30.

Marriot, R. (1981) The Christchurch Polytechnic Reading Survey, Department of Languages and Communication Studies, Christchurch Polytechnic.

Mealyea, R. (1981) 'Reading: The Role of the Trade Teacher', Research Report. Caulfield Institute of Technology, Victoria.

Nicholson, T. (1981) 'Reading Research in New Zealand'. Research in Education in New Zealand: The State of the Art. NZARE/Delta.

Parker, R.P. and Davis, F. (Eds) (1983) Language in American and British Schools: Some Responses to the Bullock Report London, Agatham.

Plumlee, L.B. (1964) 'Estimating Means and SD's from partial data - an empirical check on Lord's Item Sampling Technique', Educational and Psychological Measurement, 24 (3), 623-630.

Rankin, E. and Dale, L. (1970) 'Cloze Residual Gain - A Technique for Measuring Learning through Reading'. 18th Yearbook of the National Reading Conference. Marquette University, 74-84.

Robinson, R. (1973) 'The Cloze Procedure: A New Tool for Adult Education'. Adult Education, 23, 87-98.

Shoemaker, D.M. (1973) Principles and Procedures of Multiple Matrix Sampling. Ballinger Publishing Co., Cambridge, Mass.

Sloan, M. and Pearson, D.P. (1978) 'The Effects of Teacher Intervention Strategies on Poor Readers Comprehension of Technical Matter in a Technical-Vocational School'. The Journal of Vocational Education Research, III, 1.

Somerville, R., Chapman, E. (1979) 'Adult Reading Assistance', set, 2, item 9.

Stevens, K. (1981) 'Research on Technical Education in New Zealand: The State of the Art'. Occasional Paper, No. 2. Tutor Training Unit: Lower Hutt.

Taylor, W. (1953) 'Cloze Procedure: A New Tool for Measuring Readability'. Journalism Quarterly, 30, 415-433.

Taylor, W. (1954) 'Recent Developments in the Use of the Cloze Procedure'. Journalism Quarterly. 33, 42-48

Taylor, W. (1956) 'Cloze Readability Scores as Indices of Individual Differences in Comprehension and Aptitude'. Journal of Applied Psychology, 41, 19-26

Thornton, L.J. (1977) 'Relationship of Carpentry Material Installation Instructions and Carpentry Textbooks To Reading Ability of High School Carpentry Students'. Masters Thesis. Utica, New York: SUNY College of Technology.

Thornton, L.J. (1979) 'Reading Sensitivity in Post-secondary Occupational Education'. Lifelong Education: The Adult Years, December, 12-15 and 19.

Wagner, G.A. (1980) 'A Readability Assessment of Some Technical Correspondence Institute Assignments'. Final Report. NZCER: Wellington.

APPENDICES

I	THE CLOZE PROCEDURE	54
	1. The validity of the cloze	54
	2. Constructing a cloze test	54
	3. The interpretation and use of cloze test scores	56
	4. Generating items for the cloze test	57
	5. An example of a cloze test used in this study	59
II	THE MULTIPLE MATRIX SAMPLING PLAN	69
III	AVOIDING SERIAL EFFECTS	71
IV	DIRECTIONS FOR ADMINISTERING AND RETURNING THE CLOZE TESTS	72
V	ERROR ANALYSIS OF THE COMMON CLOZE RESPONSES	76
VI	ESSAY MARKING PROCEDURES	82
	1. The marking criteria	82
	2. Marker agreement	83
VII	RELIABILITY OF THE TEST INSTRUMENTS	84
VIII	EQUIVALENCE OF THE MMS GROUP AND THE TEST BOOKLETS	86
IX	COMPARABILITY OF THE 1982 AND 1983 SAMPLES	90
X	VALIDITY OF THE CLOZE AS A MEASURE OF EXAMINATION READABILITY	91

APPENDIX I: THE CLOZE PROCEDURE

1. *The validity of the cloze*

High correlations exist between cloze tests and other known tests of reading comprehension. Moderate to high criterion-related validity coefficients are reported by Bormuth (1967, 1968); Anderson (1969); Rankin and Dale (1970) and Robinson (1973) on standardized tests of general and specific reading skills. When the cloze and conventional multiple-choice tests are constructed from the same passages, Bormuth (1967) working with elementary children and correcting for the unreliability of the two tests employed, reported values approaching 1.00, an almost perfect correlation between the paired scores.

Further evidence on the validity of the cloze test is provided by two developmental studies by Bormuth (1969, 1971). The first showed that both cloze and multiple-choice tests of reading comprehension loaded on a single factor which accounted for 77% of the variance in test scores, thus providing evidence that cloze tests were measuring similar dimensions of reading comprehension to the more traditional multiple-choice tests. The second showed that there was a strong curvilinear relationship between cloze scores on reading passages and measures of "information gain" on the same passages. Such evidence demonstrates reasonably strong validity of the cloze test as a measure of reading comprehension.

2. *Constructing a cloze test*

A number of factors need to be considered in constructing a cloze test.

(a) *Deletion pattern*

There are two common deletion patterns: (i) every 'nth' word, and (ii) random. Both patterns are completely mechanical and can be applied by lay test constructors. The most favoured pattern is every 'nth' word. Research (eg MacGinitie, 1961; Taylor, 1956) has suggested that word intervals larger than four or five in a cloze test

ensure item independence, that is, the probability of getting an item correct was not dependent on getting the previous item correct. Traditionally, deletion rates of 5 to 10 have been adopted with no significant differences in cloze performance between the deletion rates. However, little is to be gained from a deletion rate greater than seven (Harrison, 1980). Thus, for this study, an interval of six words (deletion of every seventh word) offered the best combination of item independence and economic use of 150-170 word passages.

(b) *Type of words deleted*

The class of words to be deleted from a passage alters the nature of the cloze task. If only content words (i.e., nouns, verbs, adjectives and adverbs) are deleted from a passage the cloze test tends to become a test of prior knowledge of technically-worded content. However, to obtain a global measure of reading comprehension or text readability, systematic nth word deletion patterns in which all words are candidates for deletion are appropriate because no special emphasis is given to either semantic or syntactical factors.
In Taylor's (1957) view:

> "To restrict deletions to particular kinds of words is to ignore the fact that these kinds might not occur equally often in different materials. That difference in frequency of occurrence may itself be a readability factor; if so, its effects should be included in - not excluded from - the results." (p.25).

Further, several researchers (e.g., Elley, 1976) have demonstrated that the difficulty of the cloze task is influenced by the class of words to be replaced. In general, 'structural' words (pronouns, prepositions, conjunctions, etc) are easier to replace than 'content' words (nouns, verbs, adverbs, adjectives). Further, nouns

are substantially more difficult to replace than other content words. Elley (1976) provided evidence that the percentage of nouns, verbs, adverbs, adjectives, pronouns, conjunctions and prepositions correctly replaced in a cloze exercise was 28.6, 72.4, 52.1, 38.9, 72.6, 59.4 and 74.4, respectively. Thus, the difficulty of a cloze task is strongly influenced by the proportion of nouns deleted. As pieces of writing may vary in style, some containing many nouns and some very few, the analysis of each test passage for proportion of nouns present, compared with the proportion of nouns deleted, helped to ensure that the difficulty of the cloze task accurately reflected the difficulty of the test passage.

(c) *Number of words deleted*

To be reliable, a test must be sufficiently long. Acting on Taylor's (1965) earlier suggestions that a cloze test should contain at least fifty cloze items, Bormuth (1962) found that cloze tests of 50 items (deletions) had the same range of reliabilities as multiple-choice tests of reading comprehension containing 31 items (from .75 to .95). He also developed a table to estimate the optimum test length based upon different standard errors of measurement (Bormuth, 1956) in Hansen and Hesse (1974). The research design and the level of measurement found to be acceptable resulted in tests of four passages each of 20 blanks (making a total of 80 blanks) being constructed.

3. *The interpretation and use of cloze test scores*

In interpreting multiple-choice reading comprehension test scores it has been traditional practice to determine if the students have reached the *'instructional'* level of comprehension (i.e., 75 percent correct). Reading material at this level, or easier, is said to be at the students' *instructional* level because the students are presumed ready for supervised instruction. Passages in which readers can achieve 90 percent correct, or better, are said to be at the *independent* level, and

students reading at this level would be expected to work without direct support from the instructor or tutor. Similarly, it is accepted practice now to consider that scores below 75 percent correct show that the extract is at the readers' *frustration* level where even assistance from the instructor is of no real help in enabling the students to understand the materials. These specific percentage levels have been widely used as criteria and serve as useful reference points in dealing with individual or group cloze test performance.

The notion of 75 percent and 90 percent criterion levels has been matched to the concept of information gain by Bormuth (1971). Using regression procedures he related student performance on tests of information gain to their performance on cloze tests and found that for cloze scores of up to 35 percent there was little increase in information gain. Beyond 35 percent, the amount of information gain increased steeply until criterion scores between 50 percent and 54 percent were reached. On the basis of these studies, "35 percent would appear a defensible criterion of the threshold of literacy when applied to cloze scores of specified subjects on specific materials" and "50 percent might be posited as a criterion of comfortability, that point above which the subject will be able to gain information from the material with some ease and comfort" (Hansen and Hesse, 1974, pp.18-19).

In terms of information gain, scores on cloze tests between 0-34 percent, 35-50 percent and 51-100 percent can now be related to the frustrational, instructional and independent reading ability levels.

In addition, the average (mean) score obtained by a group of individuals on a cloze passage provides a measure of the readability of that passage for that group.

4. *Generating items for the cloze test*

The procedure adopted for generating cloze test items was as follows:

(i) A number between 8 and 14 was randomly selected (to ensure that at least seven words of the discourse always preceded the first blank and that the context of the passage was established).

(ii) That number of words into the passage was counted.

(iii) That word was underlined as the first to be deleted.

(iv) Every following seventh word was underlined until a total of 20 words had been chosen.

(v) Each passage was analysed for the following information:

(a) total word count (T)
(b) the proportion of words that were nouns $(\frac{N}{T})$
(c) the proportion of words underlined that were nouns $(\frac{Nu}{20})$.

If the proportion of nouns to be deleted equalled the proportion of nouns in the passage as a whole (i.e., if $(\frac{Nu}{20} = \frac{N}{T})$ the deletion pattern was accepted.

If this was not so, the 'clozing' procedure was repeated, using the immediately following word as the starting point for deleting. This was repeated until a deletion pattern was reached which included a number of deleted nouns proportional to the total nouns in the passage as a whole.

TEST OF READING COMPREHENSION

5. *An example of a cloze test used in this study.*

NAME: _____ _____ AGE: _____ years _____ months
(First) (Last)

INSTITUTE: _____ SEX: M/F DATE: _____

Date of Starting High School: _____ month _____ year

Date of Leaving High School: _____ month _____ year

Date of Starting Your Present Job: _____ month _____ year

Date of Starting Your Apprenticeship: _____ month _____ year

Educational Qualifications (tick)

- ☐ 2 years secondary education
- ☐ School Certificate (SC)
- ☐ Sixth Form Certificate
- ☐ University Entrance (UE)
- ☐ Other (please specify) _____

Are You a TCI Student? Yes/No

Section 1 — This Test

This is a test of how well you can read. Some words have been left out of the passages, and you are to write the missing word in each space.

Now try the following example. Just write the word you think has been left out, in the space provided.

New Zealand Forests

A large part of New Zealand _____ still covered with forests. Most of the _____ are owned by the state _____ are logged for their timber. The pine _____ is the most common species in these forests.

The words which best fill the blanks are: 'is'; 'forests'; 'and'; 'tree'.

Try to fill in every blank, even if you have to guess.

Here are some clues to help you:

1. Write only *one* word in each space.
2. Do the easy ones first and come back to the harder ones.
3. Spell the word as it sounds if you're not sure of the spelling.
4. If an ordinary word doesn't fit, try:
   (a) a number like *134* and *1979*,
   (b) words like *can't* and *isn't*,
   (c) abbreviations like *mm* and *NZ*, and
   (d) a word that goes together with other words, like *up-to-date*

You have 50 minutes to complete this section of the test.

DO NOT TURN OVER UNTIL YOU ARE TOLD.

RESUSCITATION

1. Break circuit immediately, but do not use metal _____ any moist material to free victim _____ circuit.
2. Instantly attend to the victim's _____ as soon as he is clear _____ conductor, feel rapidly with your finger _____ his mouth or throat and remove _____ foreign body (false teeth, gum, etc.).
3. _____ apply artificial respiration AT ONCE in _____ manner that you have been instructed _____, this should take place as close _____ the accident as possible.
4. As soon _____ artificial respiration has been commenced, and _____ it is being continued, an associate _____ loosen any tight clothing about subject's _____, chest, waist.
5. Continue artificial respiration WITHOUT _____ for up to 4 hours, and _____ then it should not cease until _____ is pronounced dead by a physician.
6. _____ natural breathing stops after being restored, _____ artificial respiration again.
7. Do not give _____ any liquid by mouth until fully restored.
8. After victim has recovered, treat as for shock, keep warm, keep quiet, plenty of fresh air.

E1

THE AMMETER

The ammeter is an instrument used to measure in amperes the current in a circuit. Because _____ flows along the connecting wires of _____ circuit, the circuit must be broken _____ and the meter connected in the _____ so that all the current flowing _____ the connecting wires also flows through _____ ammeter. The ammeter has therefore a _____ low resistance and is connected in a _____.

An ammeter must not be _____ directly across a source of supply _____ across a part of a circuit _____ a voltmeter is connected. If this were _____, a very high current would flow _____ the instrument and the excessive heat _____ mechanical strain produced would damage it _____ - probably beyond repair.

The moving-coil _____ and the moving-coil voltmeter operate _____ if current flows through them in _____ correct direction. Therefore, when these meters _____ connected to a circuit, the connections _____ be such that the + sign on the meter is connected to the conductor leading to the positive side of the supply.

E3

ALTERNATING-CURRENT SYSTEM

The national electrical-supply system in New Zealand is based _____ the alternating current (a.c.), which alternates _____ a frequency of 50 hertz. If _____ current (d.c.) is required, it must _____ obtained from private generating plants, from _____, or by converting the alternating current _____ direct current by a process known _____ rectification.

For some purposes, it does _____ matter if the supply is alternating _____ direct, but for others it is _____ important. Electric heaters and incandescent lamps, _____ example, will work with a.c. or _____ supplies. This is also true of _____ small universal motors used in domestic _____, for example, vacuum cleaners and cake _____. Direct current must, however, be used _____ electroplating and for battery charging.

The _____ motor has the advantage that its _____ can be altered by means of _____ control equipment, but variable speed a.c. _____ are complex in design and therefore expensive in larger sizes.

NON-LINEAR RESISTORS

Ohm's Law states - that the current flowing in a circuit is directly _____ to the voltage, and indirectly proportional _____ the resistance, at a constant temperature. _____ the temperature changes either due to _____ flowing through the circuit or by _____ in the surrounding temperature, then it _____ probable that the resistance will change _____. Copper for example, has a linear _____ resistance coefficient, where an increase in _____ produces a proportional increase in resistance. _____ some alloys show a non-linear _____ in that a change in temperature _____ not produce a proportional change in _____. Devices which make use of this _____-linear characteristic are called THERMISTORS, and _____ may have either a Positive Temperature _____ (PTC) where an increase in temperature _____ an increase in resistance; or a _____ Temperature Coefficient (NTC) where an increase _____ temperature produces a decrease in resistance.

_____ protection is one of the most direct, precise and universal methods yet developed for the protection of electrical machinery of all kinds from damage due to over-heating.

E7

LIBRARY MEMBERSHIP

All staff and students, and any interested members of the _____, are eligible to use the library _____ of charge. If you wish to _____ the library, fill in a registration _____ at the issue desk, and you _____ then receive a yellow membership card. _____ THIS SAFELY as you will be _____ to show it whenever you take _____ a book or periodical.

Plastic pockets _____ protect these cards are available at _____ desk for 10c.

BORROWING

Books are _____ for a loan period of 2 _____ 3 weeks, and can usually be _____ for a further period unless someone _____ is waiting for them. Books can _____ be renewed by telephone.

RESERVES

Any library _____ which is out on loan, may _____ reserved. It will be put aside _____ you when it is returned, and _____ will be notified that it is _____. Just ask at the issue desk for the reservation card.

C02

PUPILS WILL GET TIME OFF

The country's primary school children can expect to have an afternoon off when _____ teachers hold meetings the week after _____ to discuss the Government Service Tribunal's _____ on their salary claim.

The tribunal _____ against the teachers' salary claim for _____ pay relativity with secondary teachers, though _____ awarded some increases for school principals.

_____ deputy national secretary of the New _____ Educational Institute, Miss Helen Anderson, said _____ that no timetables for the meetings _____ yet been finalised.

Miss Anderson said _____ 92 branches of the institute would _____ hold their meetings in the afternoon _____ children would all be going to _____ on the mornings of the meetings.

"_____ have waited months to hear the _____ and we believe they're entitled to _____ out about it and to discuss _____", she said.

Miss Anderson said parents _____ be given "adequate" advance notice of _____ meetings.

"Clearly we are not going to leave the children of working parents unattended", she said.

C06

CONTRACTS OF APPRENTICESHIP

A contract of apprenticeship is an agreement between an employer, an apprentice, and _____ apprentice's guardian. The apprentice undertakes to _____ the employer for a stated time, _____ to learn the trade to which _____ is apprenticed.

The apprentice being a _____, the guardian also signs the contract _____ shares the responsibility. The employer undertakes _____ employ the apprentice for the stated _____, and to teach the apprentice his _____.

The contract has to be in _____, and duly registered with the appropriate _____ commissioner within 28 days of the _____ of the employment.

After registration one _____ of the contract is sent to _____ employer, one to the apprentice, and _____ is kept by the district commissioner.

_____ contract of apprenticeship is not something _____ be entered into lightly.

There are _____ responsibilities placed on all three of _____ signatories.

The apprentice commits himself to _____ period of service and learning that will make his future career.

COLD WATER SURVIVAL

Many scientists around the world have been studying _____ effects on humans of immersion in _____ water under conditions similar to those _____ following boating accidents. The results are _____ used to find ways to increase _____ time through various behavioural and technological _____.

Even a small increase in survival _____ can mean the difference between being _____ or dead when rescuers arrive.

Boaties (_____ others in danger of accidental immersion _____ cold water) should be aware of _____ factors that determine body cooling rate _____ eventual death from hypothermia. Such knowledge _____ improve chances of survival if an _____ occurs. Remember, drowning is a problem _____ is easily solved by use of _____ approved life jacket. Hypothermia is a _____ that is not easily solved, and _____ your careful attention.

Hypothermia means lowered, _____-body temperature. In cold water, the _____ and external tissues become cooled very rapidly. But it takes 10-15 minutes before the temperature of the heart and brain begin to cool.

SECTION 2

Write 10 to 15 lines on:

"Why I became an apprentice."

(Write as neatly and as clearly as possible.)

GO BACK AND CHECK YOUR WORK FOR SPELLING AND GRAMMAR.

APPENDIX II: THE MULTIPLE MATRIX SAMPLING PLAN

The level of statistical analysis to be performed and the level of acceptable error had to be balanced in order to determine the optimal sampling plan (i.e., the appropriate ratio of people to items). Bormuth's table (see Hansen and Hesse, 1974, p.37) of estimated standard errors for a variety of sampling plans was used to derive the sample size of about 60, with 80 items per test. The sample size of 60 test takers and the total number of apprentices available in each trade area required no more than four parallel test booklets to be constructed. One general passage and one trade passage was common to each parallel test booklet. This was done to enable checks on the equivalence of the groups to carried out.

The resulting sampling plan represented the optimal balance of the factors mentioned in Chapter 2.3 for this particular study. A summary of the MMS sampling appears in Table A.1.

TABLE A.1. *Multiple Matrix Sampling Plan*

	General	Trade
Total passages	13	7
Test takers per passage	60	120
Passages per booklet	4	4
Common passages per booklet	1	1
Number of booklet forms	4	2
Maximum time for testing	50 minutes	
(Smallest) Population available:	approximately 240	

Although approximately 60 students would respond to any one of four parallel subtests of general reading comprehension, consisting of four cloze passages, MMS analyses combine the performance of each subgroup on each subtest and estimate how the group as a whole would perform on all the test passages, that is, the level and

spread of test scores that would have been obtained if all 240 test takers had responded to all 13 test passages.

APPENDIX III: AVOIDING SERIAL EFFECTS

The order in which the sections in the test booklets were compiled was systematically reversed for half the test booklets, while the order of passages within each section remained constant. Thus, for each trade area 16 different test booklets were compiled. Figure A.1 illustrates the structure of the test booklets.

FIGURE A.1 *The Structure of the Test Booklets*

Section A (followed by)	Section B	
	Subtest B1	Subtest B2
Subtest A1	X	X
Subtest A2	X	X
Subtest A3	X	X
Subtest A4	X	X

Section B (followed by)	Section A			
	Subtest A1	Subtest A2	Subtest A3	Subtest A4
Subtest B1	X	X	X	X
Subtest B2	X	X	X	X

Each form of Section A (A1, A2, A3, A4) was combined with each form of Section B (B1, B2). Therefore, the test booklets were as follows: A1, B1; A1, B2; A2, B1; A2, B2; A3, B1; A3, B2; A4, B1; A4, B2; plus the reverse of B1, A1; B2, A1 etc.

APPENDIX IV: DIRECTIONS FOR ADMINISTERING AND RETURNING THE CLOZE TESTS

DIRECTIONS FOR ADMINISTERING THE CLOZE TESTS OF READING COMPREHENSION

Introduction

The cloze tests of reading comprehension are being administered to a number of groups of students as part of a survey of the reading abilities of "young adults". In addition, we are gathering a small sample of the student's writing. It is important that the administration of the tests be as standardized as possible for all groups so that test scores may be combined or compared. Hence, we would like to emphasize that the procedure for administering the tests, outlined below, be adhered to as closely as possible.

Administering the Test

1. Briefly explain that this is a test of reading comprehension and writing, and stress the importance of them giving their best effort during the testing period.

2. Distribute the test booklets. A sampling procedure has been used to construct the test so that there are *four* different, but parallel, test forms (A, B, C and D). The test booklets have been randomly arranged and the instructions are identical for all tests. Simply distribute the booklets to the students in the order they have been arranged.

3. Instruct the students to complete the details on the top of the front page. (Allow time for everyone to finish this.)

Directions for Section 1

4. Read aloud (clearly and distinctly):

"Follow the instructions for Section 1 on the front page while I read them aloud.

This Test

This is a test of how well you ___can___ read. Some words have been left out ___of___ the passages, and you are to write ___the___ missing word in each space.

Just write the word you think has been missed out in the space provided. Make the best guess you can. Try the following example:

New Zealand Forests

A large part of New Zealand ___(blank)___ still covered with forests. Most of the ___(blank)___ are owned by the state ___(blank)___ are logged for their timber. The pine ___(blank)___ is the most common species in these forests."

Allow a few minutes for everyone to complete the example. Then read "The words which best fill the blanks are: 'is'; 'forests'; 'and'; and 'tree'.

Try to fill in every blank, even if you have to guess.

Here are some clues to help you:

(1) Write only one word in each space.

(2) Do the easy ones first and come back to the harder ones.

(3) Spell the word as it sounds if you're not sure of the spelling.

(4) If an ordinary word doesn't fit, try:

(a) a number like 134 and 1979,
(b) words like can't and isn't,
(c) abbreviations like mm and NZ, and
(d) a word that goes together with other words, like up-to-date.

Are there any questions?" (Repeat the appropriate sections of the instructions if necessary, to clarify the procedure.)

"There are 8 passages in total. Do not forget the passages on the back of the pages.

You have 50 minutes to complete this section.

You may start now."

5. Stop them working after 50 minutes.

Directions for Section 2

6. Say *"Turn to Section 2 on Page 5 of the tset booklet. You are to write 10 to 15 lines on Why I became an apprentice. Please write as neatly and as clearly as possible. You have 10 minutes."*

7. Stop them after 10 minutes.

8. Ask them to make sure that all the details have been completed on the top of the front page and collect in the test booklets.

After the Testing

9. As soon as possible after the testing of each group of apprentices please check that all the details asked for on the front of the test booklets have been provided by the students and forward the test booklets to NZCER using the addressed labels provided. The tests will be marked by NZCER staff. It is expected that a summary of the results will be available later in the year.

If you have any queries about the test administration, please contact Alison Gilmore. Queries about the supply and return of test booklets should be directed to Felicity O'Connor. Both researchers can be contacted at:

New Zealand Council for Educational Research
P.O. Box 3237
WELLINGTON or Telephone: 847-939

Thank you very much for your participation in this survey. We much appreciate the time you have given to this task.

Graham Wagner
(Senior Research Officer)

Alison Gilmore
(Research Officer)

Felicity O'Connor
(Research Assistant)

APPENDIX V: ERROR ANALYSIS OF THE COMMON CLOZE RESPONSES

The incorrect cloze responses made by a systematic sample of the test takers on the common passages in the general and trade reading comprehension section were examined in detail with respect to their semantic and syntactic appropriateness.

A response which did not match exactly the word that had been deleted was categorized into one of the following six error types:

1 (a) *Appropriate Replacements* (AR) - meaningful replacements that were semantically and syntactically appropriate. These may be the same or a different part of speech from the correct response. The distinguishing feature is that the meaning of the sentence was preserved.

1 (b) Appropriate replacements in which more than one word (usually two) were used to replace the deleted words. The test takers were asked to write only one word in each underlined space.

2 *Grammatically Inappropriate Replacements* (GIR) - replacements which were syntactically and semantically appropriate but which were incorrect grammatically with respect to tense, number, gender or case.

3 *Partially Inappropriate Replacements* (PIR) - replacements which were syntactically acceptable but semantically unacceptable, that is, replacements which were the same part of speech as the deleted word but did not preserve the semantic sense of the sentence.

4 *Wholly Inappropriate Replacements (*WIR*)* - replacements which were both syntactically and semantically incorrect; that is, replacements which were not the same part of speech and did not retain the semantic sense of the sentence.

5 *Omissions* (O) - items to which there was no response.

Examples: In New Zealand there lives a little creature called the Kiwi.

Type of Replacement		Response
1(a)	appropriate (AR)	"small" for "little"
1(b)	appropriate (more than one word)	"known as" for "called"
2	grammatically inappropriate (GIR)	"live" for "lives"
3	partially inappropriate (PIR)	"shouted" for "called"
4	wholly inappropriate (WIR)	"some" for "lives"

The number (and proportion) of cloze errors by each type made on the common passages in the general and trade reading comprehension sections are presented in Tables A.2 and A.3 and graphically in Figure A.2.

TABLE A.2 *Mean Number ($\bar{x}$) and Proportion (p) of Errors by Type Made by Apprentices in Each Trade Area on the Common General Passages*

		Error Type						
		AR(1a)	AR(1b)	GIR(2)	PIR(3)	WIR(4)	O (5)	TOTAL
Secondary Schools								
(n = 92)	p	(0.28)	(0.01)	(0.005)	(0.37)	(0.20)	(0.14)	
	$\bar{x}$	2.45	0.10	0.04	3.20	1.73	1.20	8.71
Electrical								
(n = 59)	p	(0.34)	(0.03)	(0.01)	(0.33)	(0.14)	(0.14)	
	$\bar{x}$	2.37	0.24	0.10	2.34	0.98	1.02	7.05
Hairdressing								
(n = 65)	p	(0.26)	(0.03)	(0.004)	(0.29)	(0.17)	(0.25)	
	$\bar{x}$	2.14	0.23	0.03	2.42	1.35	2.03	8.20
Meat Retailing								
(n = 52)	p	(0.18)	(0.02)	(0.004)	(0.28)	(0.21)	(0.31)	
	$\bar{x}$	1.79	0.25	0.04	2.83	2.10	3.19	10.19
Farming								
(n = 52)	p	(0.27)	(0.02)	(0.02)	(0.26)	(0.18)	(0.25)	
	$\bar{x}$	2.40	0.17	0.13	2.37	1.63	2.25	8.96
Auto Engineering								
(n = 65)	p	(0.32)	(0.04)	(0.01)	(0.29)	(0.17)	(0.17)	
	$\bar{x}$	2.31	0.31	0.05	2.14	1.25	1.26	7.31
TOTAL								
(n = 385)	p	(0.27)	(0.02)	(0.01)	(0.31)	(0.18)	(0.21)	
	$\bar{x}$	2.27	0.21	0.06	2.59	1.51	1.73	8.37

FIGURE A.2.: *Proportion of Errors by Type Made by Each Group of Apprentices on the Common Passages.*

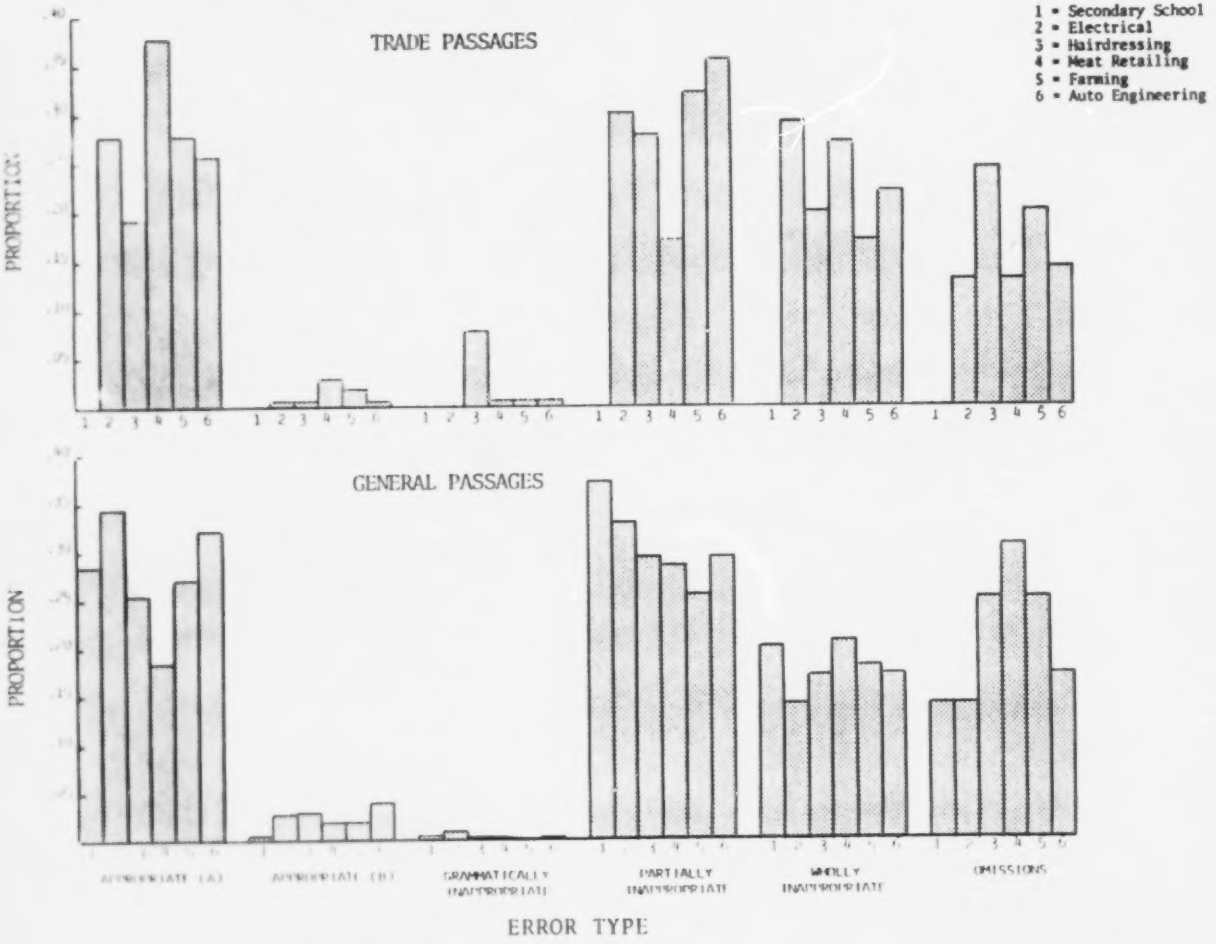


TABLE A.3 *Mean Number ($\bar{x}$) and Proportion (p) of Errors by Type Made by Apprentices in Each Trade Area on the Common Trade Passages*

		Error Type AR(1a)	AR(1b)	GIR(2)	PIR(3)	WIR(4)	O (5)	TOTAL
Electrical								
(n = 60)	p	(0.27)	(0.01)	(0.00)	(0.30)	(0.29)	(0.13)	
	$\bar{x}$	2.25	0.10	-	2.48	2.43	1.13	8.40
Hairdressing								
(n = 65)	p	(0.19)	(0.01)	(0.08)	(0.28)	(0.20)	(0.24)	
	$\bar{x}$	1.74	0.09	0.74	2.62	1.97	2.23	9.38
Meat Retailing								
(n = 52)	p	(0.38)	(0.03)	(0.01)	(0.17)	(0.27)	(0.13)	
	$\bar{x}$	3.52	0.33	0.08	1.63	2.54	1.25	9.35
Farming								
(n = 56)	p	(0.28)	(0.02)	(0.01)	(0.32)	(0.17)	(0.20)	
	$\bar{x}$	2.82	0.20	0.05	3.16	1.71	1.96	9.91
Auto Engineering								
(n = 69)	p	(0.26)	(0.01)	(0.01)	(0.36)	(0.22)	(0.14)	
	$\bar{x}$	2.55	0.07	0.12	3.61	2.23	1.36	9.94
TOTAL								
(n = 302)	p	(0.27)	(0.02)	(0.02)	(0.29)	(0.23)	(0.17)	
	$\bar{x}$	2.53	0.15	0.21	2.75	2.17	1.60	9.40

The General Reading Passages

As might be expected, the relative mean number of errors made by each apprentice group is the converse of their mean cloze test scores reported in earlier sections. Electrical apprentices made the least number of errors ($\bar{x}$ = 7.05) while Meat Retailing apprentices made the greatest number ($\bar{x}$ = 10.19).

A reasonably large number of the 'errors' made by apprentices on this cloze passage were actually synonyms for the deleted word, that is, they were appropriate replacements. From 18 percent to 34 percent of the responses were of this type with an additional 2 percent being multi-word synonyms. Some cloze marking procedures score synonyms as correct. However, to avoid marker subjectivity as to which responses constitute acceptable synonyms and which do not, standard cloze marking procedures which accept exact replacements only are normally adopted. In addition to the very

high relationship between the two schemes of scoring, marks derived from exact replacements are currently more readily interpretable with respect to established criterion cut-off scores. The Electrical and Auto Engineering apprentices made a greater number of responses which were semantically and syntactically acceptable and which preserved the meaning of the sentence.

An equally large proportion of errors were partially appropriate, that is, they were syntactically acceptable but did not preserve the semantic sense of the sentence. The group of secondary school students made the greatest number of this type of error ($\bar{x}$ = 3.20) while those for the five trade groups were lower (approximately 2.4). Approximately 20 percent of the errors were wholly inappropriate.

A quite variable pattern of omissions was found over the six groups with 31 percent of the errors made by Meat Retailing apprentices resulting from no attempt being made on the items. Hairdressing and Farming apprentices both had 25 percent of their errors of this type while for the remaining three groups the proportions were substantially lower, approximately 14 percent.

The classification of errors may be considered as a hierarchy of seriousness. For example, synonyms (type 1) are acceptable on both semantic and syntactic considerations, while type 3 replacements are inappropriate on one of these two dimensions and type 4 errors are inappropriate on both dimensions.

In considering the relative 'seriousness' of the type of errors made, the Electrical and Auto Engineering apprentices made fewer errors in total which were of a less 'serious' nature. That is, the incorrect responses tended to be appropriate or partially inappropriate rather than wholly inappropriate or omissions. In comparison, however, Meat Retailing apprentices tended to make the greatest number of errors which were generally of the more serious type, or were omissions. The remaining three groups displayed somewhat similar overall error patterns, with the greater number of errors made by secondary school students that were partially inappropriate (37 percent) being counterbalanced by the Hairdressing and Farming apprentices making more omissions (25 percent).

A Note on the Trade Reading Passages

No direct comparisons in error patterns between apprentice groups on the common trade passages are possible because each trade responded to a different cloze trade passage. However, the overall pattern of error types on the trade passages is reasonably similar to that on the general passages. Disregarding the secondary school students, the comparable proportions of errors by types 1 to 5 for the general and trade passages are 0.27, 0.03, 0.01, 0.29, 0.17, 0.23 and 0.27, 0.02, 0.02, 0.29, 0.23, 0.17, respectively.

Fluctuations in the proportions of error types for the passages from the two reading comprehension sections are more likely due to differences in the passages themselves rather than to differences arising from distinctions between trade and general reading materials generally.

APPENDIX VI: ESSAY MARKING PROCEDURES

1. The Marking Criteria

Suggested criteria for assessing 'Why I became an apprentice', based on ten sample essays scored from 1 to 5. As the writing exercise is a brief one (10 to 15 lines), a short scale seemed most appropriate. Number of markers should be at least two; if they have a substantial measure of agreement a third marker need only be called upon for widely diverging responses.

General Criteria

Note: 1 is a low score, 5 a high score. 3 is not to be thought of as a pass, but rather a mid-point on a continuum.

1-2: Essays awarded 1 or 2 have clear inadequacies, such as very brief remarks which make little attempt at answering the topic, i.e., why they became an apprentice, or have a large number of language usage errors, or are poorly argued (merely stating reasons without discussing them).

3: Generally 3s are adequate responses in that the reasons given are discussed sensibly; however, a range of mistakes and oddities of style or presentation may be present. Remember that longer essays will tend to have more errors, so some allowance has to be made for this.

4-5: Clearly written essays, with at least two sound reasons discussed at reasonable length and in a coherent fashion. A small number of errors may be present in which case a 4 is awarded; 5s are for error-free answers, although some crossing-out is permitted.

2. Marker Agreement

Once the marking criteria had been established, an agreement trial of the two markers was carried out on a sample of 34 scripts so that the degree of consistency in interpretation of the criteria could be checked before the marking proper began. A Pearson product-moment correlation of 0.79 was found between the scores awarded by the two markers with one marker assigning lower marks overall than the other (see Table A.4). A discussion between the markers of specific decisions made in awarding marks to individual scripts clarified the criteria for both.

For measuring the intra-rater reliability for both markers, the scripts marked in the trial were remarked part way through marking of the total of 1361 scripts. The original (trial) scores had been removed. The level of agreement obtained is also summarised in Table A.4.

TABLE A.4 *Marker Agreement in Assigning Scores to Scripts (n = 34)*

		Trial Marking	Re-marking	Intra-marker r
Marker 1	$\bar{X}$	3.18	3.41	0.83
	S	0.98	0.86	
Marker 2	$\bar{X}$	3.35	3.00	0.80
	S	0.85	0.78	
Inter-marker r*		0.79		

*r corrected for coarseness of the scale

A retrospective check on the consistency between markers achieved in assigning grades to all 1361 scripts showed that the median correlation (over trade groups) was 0.82 (corrected). The levels of inter-rater and intra-rater reliability in marking were considered to be acceptably high for this exercise.

APPENDIX VII: RELIABILITY OF THE TEST INSTRUMENTS

Part I

Before examining the results in detail a check was made of the degree of precision (or reliability) of the three components of the test instrument the scores of which were analysed independently, namely (i) the 'General' section of reading comprehension; (ii) the 'trade' section of reading comprehension; and (iii) the essay writing sections.

Estimates of reliability (internal consistency) of the reading comprehension sections were obtained by applying the Kuder-Richardson 21 formula to the estimated test parameters from the MMS design, in which the total numbers of cloze 'items' for the general and trade sections were 320 and 160, respectively. The resulting reliability coefficients are reported below:

TABLE A.5 *KR-21 Estimates of Reliability of Cloze Reading Comprehension Tests*

Group	General Section	Trade Section
Secondary Schools	0.95	-
Electrical	0.94	0.91
Hairdressing	0.95	0.92
Meat Retailing	0.95	0.92
Farming	0.97	0.93
Auto Engineering	0.96	0.93

Clearly, the reliability coefficients obtained were very high indeed (greater than 0.94 for the general section and greater than 0.91 for the trade sections) indicating high levels of internal consistency within these two test components.

As reported in the previous appendix, checks on the intra-rater and inter-rater reliability of the essay marking were 0.82, 0.80 and 0.79, respectively; satisfactory levels of marker consistency.

Part II

A check was made on the reliability of that part of the test used to provide a measure of examination readability - the examination extracts - by using the Kuder-Richardson 21 formula. The reliability coefficients are as follows:

TABLE A.6 *Reliability (KR-21) of Cloze Examination Tests*

Group	r
Electrical	.88
Auto Engineering	.90
Hairdressing	.82
Meat Retailing	.97

APPENDIX VIII: EQUIVALENCE OF THE MMS GROUPS AND THE TEST BOOKLETS

As already outlined, the sampling design employed in this study involved constructing parallel test forms (four for the general reading comprehension section, and two for the trade reading comprehension section). The procedures used, firstly to assign reading passages randomly to test booklets and, secondly, to distribute the test booklets randomly to test takers, were taken to help ensure parallelism of the subtests in terms of difficulty and equivalence of the apprentice groups answering each subtest. A check on the equivalence achieved in these two respects was made before subsequent analyses were undertaken.

(a) Equivalence of the Groups Taking Parallel Test Forms

To enable a check to be made of the equivalence in ability of the apprentices within each trade group answering each of the different test forms, one reading passage was made common to each test (for both the general and trade reading sections).

Table A.7 summarises the means and standard deviations on the common general reading passage for each test form and apprentice group. It is clear that for each trade area the mean scores for each group are very similar, with the greatest difference between any two groups being within half a standard deviation. Statistical analyses revealed no significant differences between the four groups within each trade except for Hairdressing ($p < 0.05$).

TABLE A.7 *Equivalence of MMS Groups - General: Means and Standard Deviations of Groups Within Each Trade Area on the Common General Passage*

GROUP	SUBTEST							
	A		B		C		D	
	$\bar{X}$	S	$\bar{X}$	S	$\bar{X}$	S	$\bar{X}$	S
Secondary Schools	11.18	3.12	11.67	3.12	11.75	3.26	11.63	3.55
Electrical	12.99	2.24	13.07	3.00	12.46	3.39	12.21	2.70
Hairdressing	11.30	3.01	10.42	3.66	12.17	3.54	10.78	3.95
Meat Retailing	11.00	3.42	11.00	3.76	10.26	3.91	10.42	3.93
Farming	10.46	4.05	10.60	3.61	10.74	3.90	10.29	3.78
Auto Engineering	11.75	3.24	11.81	3.47	12.05	3.35	11.87	2.39

In a similar manner, performance on the common trade passage by the two different groups to whom the two parallel tests were administered was examined. No significant differences were found in the mean scores between the pairs of groups of any trade area. The means and standard deviations are reported in Table A.8.

These figures confirm that the procedures adopted for randomly distributing test booklets to test takers were successful and that, generally, the groups of apprentices who responded to the different subtests were equivalent in terms of their reading comprehension abilities.

TABLE A.8 *Equivalence of MMS Groups - Trade: Means and Standard Deviations of Groups Within Each Trade Area on the Common Trade Passages*

GROUP	SUBTEST			
	A		B	
	$\bar{X}$	S	$\bar{X}$	S
Electrical	11.21	2.92	11.23	2.76
Hairdressing	10.47	2.82	10.63	2.84
Meat Retailing	11.18	3.00	11.43	3.04
Farming	9.68	2.98	10.10	2.67
Auto Engineering	10.34	3.42	9.98	3.62

(b) *Parallelism of the Test Booklets*

Having established that the apprentice groups within each trade taking the different test booklets were equivalent in almost all cases, the 'parallelism' of the test booklets, with respect to difficulty, was examined. While random assignment of reading passages to test booklets assists the construction of parallel tests, because of the reasonably small number of passages involved (13 for the General Section; 7 for the Trade Sections), close parallelism was not anticipated. As Tables A.9 and A.10 demonstrate, the mean scores on the two sections of each booklet differed somewhat. A consistent pattern, however, emerged for the relative difficulty of each test booklet over trade areas. For the general section, test booklet A appeared to be least difficult while test booklet D was the most difficult.

The two trade sections for Meat Retailing, Farming and Auto Engineering trades were found to be comparable in level of difficulty, while for the remaining two groups, Electrical and Hairdressing, some statistical difference in test difficulty was found ($p < 0.01$).

The departure from strict parallelism of test booklets in each section is not a threat to further analysis, however, as MMS

computational procedures take account of such differences, while other analyses are undertaken separately for each test form.

TABLE A.9 *Parallelism of the Test Booklets - General Section: Means and Standard Deviations of Groups Within Each Trade Area on the General Section*

GROUP	SUBTEST							
	A		B		C		D	
	$\bar{X}$	S	$\bar{X}$	S	$\bar{X}$	S	$\bar{X}$	S
Secondary Schools	46.67	8.43	38.80	9.58	41.15	9.51	37.47	10.07
Electrical	49.08	6.77	44.55	8.20	43.17	8.14	41.68	8.07
Hairdressing	44.82	8.31	36.43	9.90	42.04	8.47	36.59	9.25
Meat Retailing	42.58	9.51	35.87	10.58	39.88	9.80	33.64	11.44
Farming	46.04	8.40	37.71	9.14	39.26	10.05	37.01	9.70
Auto Engineering	47.07	9.43	39.84	9.92	41.79	8.02	38.90	9.74

TABLE A.10 *Parallelism of the Test Booklets - Trade Section: Means and Standard Deviations of Groups Within Each Trade Area on the Trade Section*

GROUP	SUBTEST			
	A		B	
	$\bar{X}$	S	$\bar{X}$	S
Electrical	36.71	9.07	44.37	9.37
Hairdressing	45.31	9.53	40.21	9.45
Meat Retailing	37.49	8.93	37.78	8.40
Farming	35.95	9.22	35.80	8.66
Auto Engineering	38.90	10.19	38.35	7.99

APPENDIX IX : COMPARABILITY OF THE 1982 AND 1983 SAMPLES

TABLE A.11 *Comparability of the 1982 and 1983 Samples on the Common General and Trade Passages*

TRADE	1982			1983		
	$\bar{X}$	S	N	$\bar{X}$	S	N
Electrical General	12.7	2.8	239	12.9	2.6	39
Trade	11.2	2.8	241	11.5	3.0	39
Auto Engineering General	11.9	3.4	320	12.3	3.3	32
Trade	10.2	3.5	333	8.3	3.6	32
Hairdressing General	11.2	3.5	316	11.4	4.0	35
Trade	10.6	2.8	319	11.2	3.0	35
Meat Retailing General	10.7	3.8	205	9.9	4.0	43
Trade	11.3	3.0	208	11.2	2.4	43

APPENDIX X: VALIDITY OF THE CLOZE AS A MEASURE OF EXAMINATION READABILITY

The utility of the cloze both as a measure of reading comprehension and readability has been well documented. However, the cloze has not been used as a measure of the readability of examinations and therefore a validity check using some well-known readability formulas was undertaken. From studies comparing readability indices reported by Harrison (1980), the Mugford Charts and Forcast formula appeared most appropriate for assessing trade reading. Estimates of the readability of the trade examinations were therefore gained using these two procedures and the well-known Noun Frequency Method (Elley, 1975) and were compared with estimates obtained with the cloze procedure.

Median correlation coefficients (rank order) between readability measures within each trade on the trade and three examination extracts were obtained (the General passage was not included as it was the readability of trades materials which was being considered). These coefficients are recorded in Table A.12.

TABLE A.12 *Median Correlation Coefficients Between Readability Indices*

	Cloze	Mugford	NFM	Forcast
Cloze	-	.69	.05	.54
Mugford		-	.25	.79
NFM			-	.58
Forcast				-

Elley (1980) makes special mention of the limitations of the Noun Frequency method in dealing with any material having a specialised vocabulary and the negligible relationship with the cloze test (0.05) would appear to bear out Elley's assertion that it may not be an appropriate measure to use, without modification,

in this setting. Nor is it closely related to the Mugford index.

In view of recent research on the high validity of the Mugford Charts age level predictions (Harrison, 1980) the cloze procedure appeared to achieve a satisfactory level of concurrent validity, demonstrated by a correlation of 0.69. This confirms its application as a reasonable measure for predicting the readability of examination papers. Further confirmation of the validity of the cloze procedure, in the context of written examinations, is contained in Table A.13.

TABLE A.13 *Product-Moment Correlations*[1] *Between Scores on the Trade Cloze Passages and Examination Extracts*

	Correlation with Trade Cloze Passage			
Exam Extract	Auto Engineering (n = 32)	Hairdressing (n = 35)	Electrical (n = 39)	Meat Retailing (n = 43)
Exam 1	.57	.60	.47	.48
Exam 2	.58	1.00	.74	.43
Exam 3	.57	1.00	.70	.80
Exam Total	.57	1.00	.65	.47

all r's statistically significant $p<.05$.

[1] corrected for attenuation

When corrected for unreliability, the examination extracts corelate consistently at a statistically significant level with the trade cloze passage indicating that both the examination and trade cloze passages are measuring the same thing, that is, readability.

Given these two pieces of evidence, it may be concluded with some confidence, that the cloze procedure is an appropriate and valid method for assessing the readability of TCB examinations.